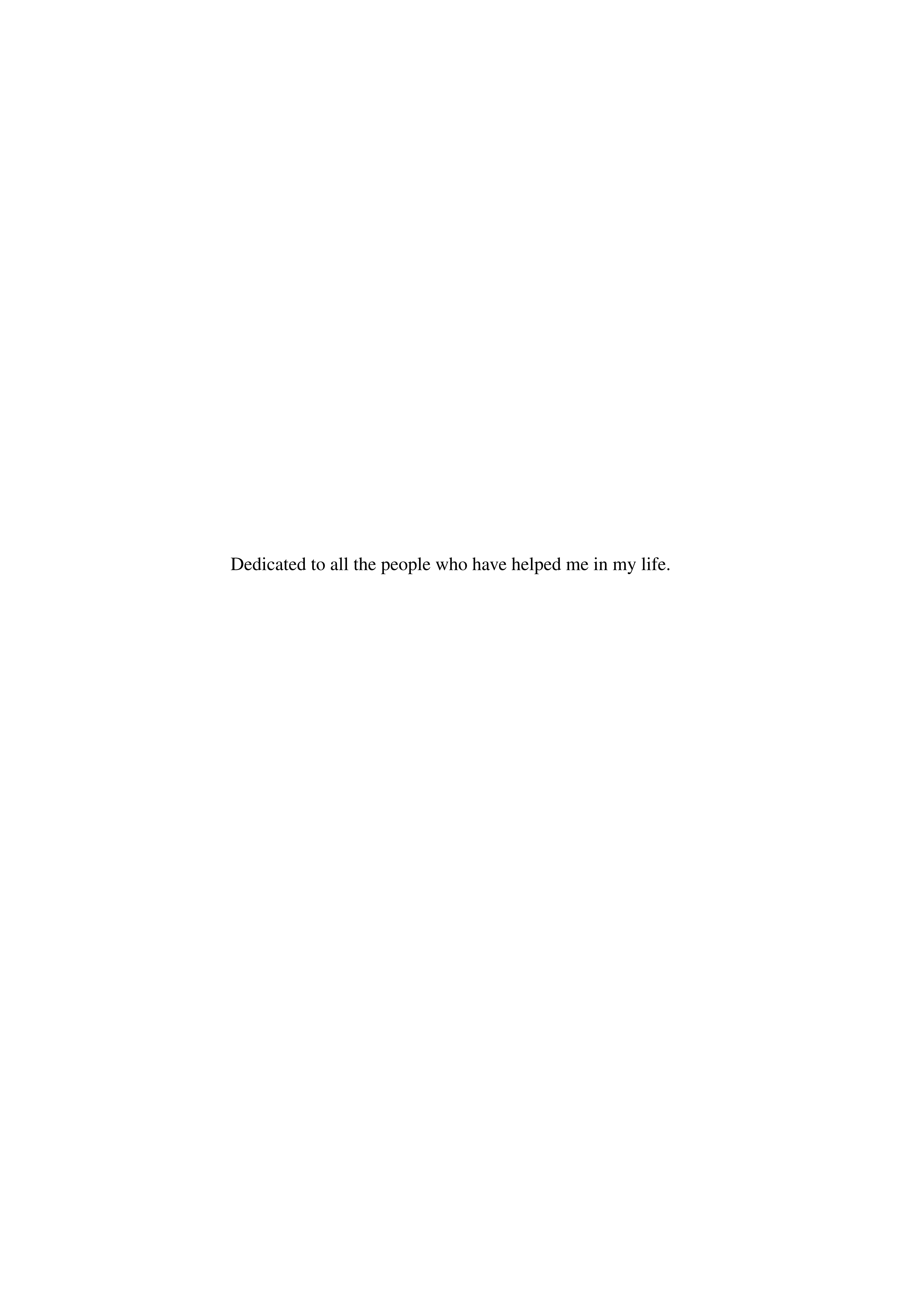
Understanding
Basic Calculus

S.K. Chung



Dedicated to all the people who have helped me in my life.

Preface

This book is a revised and expanded version of the lecture notes for *Basic Calculus* and other similar courses offered by the Department of Mathematics, University of Hong Kong, from the first semester of the academic year 1998-1999 through the second semester of 2006-2007. It can be used as a textbook or a reference book for an introductory course on one variable calculus.

In this book, much emphasis is put on explanations of concepts and solutions to examples. By reading the book carefully, students should be able to understand the concepts introduced and know how to answer questions with justification. At the end of each section (except the last few), there is an exercise. Students are advised to do as many questions as possible. Most of the exercises are simple drills. Such exercises may not help students understand the concepts; however, without practices, students may find it difficult to continue reading the subsequent sections.

Chapter 0 is written for students who have forgotten the materials that they have learnt for HKCEE Mathematics. Students who are familiar with the materials may skip this chapter.

Chapter 1 is on sets, real numbers and inequalities. Since the concept of sets is new to most students, detail explanations and elaborations are given. For the real number system, notations and terminologies that will be used in the rest of the book are introduced. For solving polynomial inequalities, the method will be used later when we consider where a function is increasing or decreasing as well as where a function is convex or concave. Students should note that there is a shortcut for solving inequalities, using the Intermediate Value Theorem discussed in Chapter 3.

Chapter 2 is on functions and graphs. Some materials are covered by HKCEE Mathematics. New concepts introduced include domain and range (which are fundamental concepts related to functions); composition of functions (which will be needed when we consider the Chain Rule for differentiation) and inverse functions (which will be needed when we consider exponential functions and logarithmic functions).

In Chapter 3, intuitive idea of limit is introduced. Limit is a fundamental concept in calculus. It is used when we consider differentiation (to define derivatives) and integration (to define definite integrals). There are many types of limits. Students should notice that their definitions are similar. To help students understand such similarities, a summary is given at the end of the section on two-sided limits. The section of continuous functions is rather conceptual. Students should understand the statements of the Intermediate Value Theorem (several versions) and the Extreme Value Theorem.

In Chapters 4 and 5, basic concepts and applications of differentiation are discussed. Students who know how to work on limits of functions at a point should be able to apply definition to find derivatives of “simple” functions. For more complicated ones (polynomial and rational functions), students are advised not to use definition; instead, they can use rules for differentiation. For application to curve sketching, related concepts like critical numbers, local extremizers, convex or concave functions etc. are introduced. There are many easily confused terminologies. Students should distinguish whether a concept or terminology is related to a function, to the x-coordinate of a point or to a point in the coordinate plane. For applied extremum problems, students

should note that the questions ask for global extremum. In most of the examples for such problems, more than one solutions are given.

In Chapter 6, basic concepts and applications of integration are discussed. We use limit of sums in a specific form to define the definite integral of a continuous function over a closed and bounded interval. This is to make the definition easier to handle (compared with the more subtle concept of "limit" of Riemann sums). Since definite integrals work on closed intervals and indefinite integrals work on open intervals, we give different definitions for primitives and antiderivatives. Students should notice how we can obtain antiderivatives from primitives and vice versa. The Fundamental Theorem of Calculus (several versions) tells that differentiation and integration are reverse process of each other. Using rules for integration, students should be able to find indefinite integrals of polynomials as well as to evaluate definite integrals of polynomials over closed and bounded intervals.

Chapters 7 and 8 give more formulas for differentiation. More specifically, formulas for the derivatives of the sine, cosine and tangent functions as well as that of the logarithmic and exponential functions are given. For that, revision of properties of the functions together with relevant limit results are discussed.

Chapter 9 is on the Chain Rule which is the most important rule for differentiation. To make the rule easier to handle, formulas obtained from combining the rule with simple differentiation formulas are given. Students should notice that the Chain Rule is used in the process of logarithmic differentiation as well as that of implicit differentiation. To close the discussion on differentiation, more examples on curve sketching and applied extremum problems are given.

Chapter 10 is on formulas and techniques of integration. First, a list of formulas for integration is given. Students should notice that they are obtained from the corresponding formulas for differentiation. Next, several techniques of integration are discussed. The substitution method for integration corresponds to the Chain Rule for differentiation. Since the method is used very often, detail discussions are given. The method of Integration by Parts corresponds to the Product Rule for differentiation. For integration of rational functions, only some special cases are discussed. Complete discussion for the general case is rather complicated. Since Integration by Parts and integration of rational functions are not covered in the course *Basic Calculus*, the discussion on these two techniques are brief and exercises are not given. Students who want to know more about techniques of integration may consult other books on calculus. To close the discussion on integration, application of definite integrals to probability (which is a vast field in mathematics) is given.

Students should bear in mind that the main purpose of learning calculus is not just knowing how to perform differentiation and integration but also knowing how to apply differentiation and integration to solve problems. For that, one must understand the concepts. To perform calculation, we can use calculators or computer softwares, like *Mathematica*, *Maple* or *Matlab*. Accompanying the pdf file of this book is a set of Mathematica notebook files (with extension .nb, one for each chapter) which give the answers to most of the questions in the exercises. Information on how to read the notebook files as well as trial version of Mathematica can be found at http://www.wolfram.com.

Contents

Chapter 0

Revision

0.1 Exponents

Definition

(1) Let n be a positive integer and let a be a real number. We define a^n to be the real number given by

$$a^n = \underbrace{a \cdot a \cdots a}_{n \text{ factors}}.$$

(2) Let n be a negative integer n, that is, $n = -k$ where k is a positive integer, and let a be a real number different from 0. We define a^{-k} to be the real number given by

$$a^{-k} = \frac{1}{a^k}.$$

(3) (i) Let a be a real number different from 0. We define $a^0 = 1$.

(ii) We do not define 0^0 (thus the notation 0^0 is meaningless).

Terminology In the notation a^n, the numbers n and a are called the *exponent* and *base* respectively.

Rules for Exponents Let a and b be real numbers and let m and n be integers (when $a = 0$ or $b = 0$, we have to add the condition: m, n different from 0). Then we have

(1) $a^m a^n = a^{m+n}$

(2) $\dfrac{a^m}{a^n} = a^{m-n}$ provided that $a \neq 0$

(3) $(a^m)^n = a^{mn}$

(4) $(ab)^n = a^n b^n$

(5) $\left(\dfrac{a}{b}\right)^n = \dfrac{a^n}{b^n}$ provided that $b \neq 0$

Exercise 0.1

1. Simplify the following; give your answers without negative exponents.

(a) $x^6 x^{-3}$ (b) $\dfrac{x^{-1}y^2}{z^{-3}}$

(c) $\left(x^{-2}y^3\right)^4$ (d) $\left(2x^2\right)^{-3} y^4 \div \left(x^{-1}y\right)^2$

0.2 Algebraic Identities and Algebraic Expressions

Identities Let a and b be real numbers. Then we have

(1) $(a+b)^2 = a^2 + 2ab + b^2$

(2) $(a-b)^2 = a^2 - 2ab + b^2$

(3) $(a+b)(a-b) = a^2 - b^2$

Remark The above equalities are called *identities* because they are valid for all real numbers a and b.

Caution In general, $(a+b)^2 \neq a^2 + b^2$. Note: $(a+b)^2 = a^2 + b^2$ if and only if $a = 0$ or $b = 0$.

Example Expand the following:

(1) $\left(\sqrt{x}+2\right)^2$

(2) $\left(x - \dfrac{5}{x}\right)^2$

(3) $\left(\sqrt{x^2+1}+7\right)\left(\sqrt{x^2+1}-7\right)$

Solution

(1)
$$\begin{aligned}\left(\sqrt{x}+2\right)^2 &= \left(\sqrt{x}\right)^2 + 2\left(\sqrt{x}\right)(2) + 2^2 \\ &= x + 4\sqrt{x} + 4\end{aligned}$$

(2)
$$\begin{aligned}\left(x - \frac{5}{x}\right)^2 &= x^2 - 2(x)\left(\frac{5}{x}\right) + \left(\frac{5}{x}\right)^2 \\ &= x^2 - 10 + \frac{25}{x^2}\end{aligned}$$

(3)
$$\begin{aligned}\left(\sqrt{x^2+1}+7\right)\left(\sqrt{x^2+1}-7\right) &= \left(\sqrt{x^2+1}\right)^2 - 7^2 \\ &= \left(x^2+1\right) - 49 \\ &= x^2 - 48\end{aligned}$$

□

Example Simplify the following:

(1) $\dfrac{x^2 - x - 6}{x^2 - 6x + 9}$

(2) $\dfrac{x^2}{x^2-1} - 1$

(3) $\dfrac{2}{x^2+2x+1} - \dfrac{1}{x^2 - x - 2}$

(4) $\left(x - y^{-1}\right)^{-1}$

(5) $\dfrac{3 + \frac{6}{x}}{x + \frac{x}{x+1}}$

Solution

(1)
$$\frac{x^2-x-6}{x^2-6x+9} = \frac{(x-3)(x+2)}{(x-3)^2} = \frac{x+2}{x-3}$$

(2)
$$\frac{x^2}{x^2-1}-1 = \frac{x^2-(x^2-1)}{x^2-1} = \frac{1}{x^2-1}$$

(3)
$$\frac{2}{x^2+2x+1}-\frac{1}{x^2-x-2} = \frac{2}{(x+1)^2}-\frac{1}{(x+1)(x-2)} = \frac{2(x-2)-(x+1)}{(x+1)^2(x-2)} = \frac{x-5}{(x+1)^2(x-2)}$$

(4)
$$\left(x-y^{-1}\right)^{-1} = \left(x-\frac{1}{y}\right)^{-1} = \left(\frac{xy-1}{y}\right)^{-1} = \frac{y}{xy-1}$$

(5)
$$\frac{3+\frac{6}{x}}{x+\frac{x}{x+1}} = \frac{\frac{3x+6}{x}}{\frac{x(x+1)+x}{x+1}} = \frac{\frac{3x+6}{x}}{\frac{x^2+2x}{x+1}} = \frac{3(x+2)}{x}\cdot\frac{x+1}{x(x+2)} = \frac{3(x+1)}{x^2}$$

□

FAQ What is expected if we are asked to simplify an expression? For example, in (5), can we give $\frac{3x+3}{x^2}$ as the answer?

Answer There is no definite rule to tell which expression is simpler. For (5), both $\frac{3(x+1)}{x^2}$ and $\frac{3x+3}{x^2}$ are acceptable. Use your own judgment. □

Exercise 0.2

1. Expand the following:

(a) $(2x+3)^2$ (b) $(3x-y)^2$

(c) $(x+3y)(x-3y)$ (d) $(x+3y)(x+4y)$

(e) $\left(2\sqrt{x}-3\right)^2$ (f) $\left(\sqrt{x}+5\right)\left(\sqrt{x}-5\right)$

2. Factorize the following:

(a) $x^2-7x+12$ (b) x^2+x-6

(c) $x^2+8x+16$ (d) $9x^2+9x+2$

(e) $9x^2-6x+1$ (f) $5x^2-5$

(g) $3x^2-18x+27$ (h) $2x^2-12x+16$

3. Simplify the following:

(a) $\dfrac{x^2-x-6}{x^2-7x+12}$ (b) $\dfrac{x^2+3x-4}{2-x-x^2}$

(c) $\dfrac{2x}{x^2-1}\div\dfrac{4x^2+4x}{x-1}$ (d) $\dfrac{\frac{1}{x+h}-\frac{1}{x}}{h}$

0.3 Solving Linear Equations

A *linear equation* in one (real) unknown x is an equation that can be written in the form

$$ax+b=0,$$

where a and b are constants with $a \neq 0$ (in this course, we consider real numbers only; thus a "*constant*" means a real number that is fixed or given). More generally, an equation in one unknown x is an equation that can be written in the form

$$F(x)=0 \tag{0.3.1}$$

Remark To be more precise, F should be a function from a subset of $\mathbb{R}$ into $\mathbb{R}$. See later chapters for the meanings of "*function*" and "$\mathbb{R}$".

Definition A *solution* to Equation (0.3.1) is a real number x_0 such that $F(x_0)=0$.

Example The equation $2x+3=0$ has exactly one solution, namely $-\frac{3}{2}$.

To solve an equation (in one unknown) means to find all solutions to the equation.

Definition We say that two equations are *equivalent* if the have the same solution(s).

Example The following two equations are equivalent:

(1) $2x+3=0$

(2) $2x=-3$

To solve an equation, we use properties of real numbers to transform the given equation to equivalent ones until we obtain an equation whose solutions can be found easily.

Properties of real numbers Let a, b and c be real numbers. Then we have

(1) $a = b \Longleftrightarrow a + c = b + c$

(2) $a = b \Longrightarrow ac = bc$ and $ac = bc \Longrightarrow a = b$ if $c \neq 0$

Remark

- $\Longrightarrow$ is the symbol for "*implies*". The first part of Property (2) means that if $a = b$, then $ac = bc$.
- $\Longleftrightarrow$ is the symbol for "$\Longrightarrow$ and $\Longleftarrow$". Property (1) means that if $a = b$, then $a + c = b + c$ and vice versa, that is, $a = b$ iff $a + c = b + c$. In mathematics, we use the shorthand "*iff*" to stand for "*if and only if*".

Example Solve the following equations for x.

(1) $3x - 5 = 2(7 - x)$

(2) $a(b + x) = c - dx$, where a, b, c and d are real numbers with $a + d \neq 0$.

Solution

(1) Using properties of real numbers, we get

$$\begin{aligned} 3x - 5 &= 2(7 - x) \\ 3x - 5 &= 14 - 2x \\ 3x + 2x &= 14 + 5 \\ 5x &= 19 \\ x &= \frac{19}{5}. \end{aligned}$$

The solution is $\frac{19}{5}$.

FAQ Can we omit the last sentence?

Answer The steps above means that a real number x satisfies $3x - 5 = 2(7 - x)$ if and only if $x = \frac{19}{5}$. It's alright if you stop at the last line in the equation array because it tells that given equation has one and only one solution, namely $\frac{19}{5}$. □

FAQ What is the difference between the word "*solution*" after the question and the word "solution" in the last sentence?

Answer They refer to different things. The first "*solution*" is solution (answer) to the problem (how to solve the problem) whereas the second "*solution*" means solution to the given equation. Sometimes, an equation may have no solution, for example, $x^2 + 1 = 0$ but the procedures (explanations) to get this information is a solution to the problem. □

FAQ Can we use other symbols for the unknown?

Answer In the given equation, if x is replaced by another symbol, for example, t, we get the equation $3t - 5 = 2(7 - t)$ in one unknown t. Solution to this equation is also $\frac{19}{5}$. In writing an equation, the symbol

for the unknown is not important. However, if the unknown is expressed in t, all the intermediate steps should use t as unknown:

$$\begin{aligned} 3t-5 &= 2(7-t) \\ &\vdots \\ t &= \frac{19}{5} \end{aligned}$$

□

(2) Using properties of real numbers, we get

$$\begin{aligned} a(b+x) &= c-dx \\ ab+ax &= c-dx \\ ax+dx &= c-ab \\ (a+d)x &= c-ab \\ x &= \frac{c-ab}{a+d}. \end{aligned}$$

□

Exercise 0.3

1. Solve the following equations for x.

   (a) $2(x+4) = 7x+2$ (b) $\frac{5x+3}{2} - 5 = \frac{5x-4}{4}$

   (c) $(a+b)x + x^2 = (x+b)^2$ (d) $\frac{x}{a} - \frac{x}{b} = c$

   where a, b and c are constants with $a \neq b$.

0.4 Solving Quadratic Equations

A *quadratic equation* (in one unknown) is an equation that can be written in the form

$$ax^2 + bx + c = 0 \tag{0.4.1}$$

where a, b, and c are constants and $a \neq 0$. To solve (0.4.1), we can use the *Factorization Method* or the *Quadratic Formula*.

Factorization Method The method makes use of the following result on product of real numbers:

Fact Let a and b be real numbers. Then we have

$$ab = 0 \iff a = 0 \text{ or } b = 0.$$

Example Solve $x^2 + 2x - 15 = 0$.

Solution Factorizing the left side, we obtain

$$(x+5)(x-3) = 0.$$

Thus $x+5 = 0$ or $x-3 = 0$. Hence $x = -5$ or $x = 3$. □

FAQ Can we write "$x = -5$ and $x = 3$"?

Answer The logic in solving the above equation is as follows

$$\begin{aligned} x^2 + 2x - 15 = 0 &\iff (x+5)(x-3) = 0 \\ &\iff x + 5 = 0 \text{ or } x - 3 = 0 \\ &\iff x = -5 \text{ or } x = 3 \end{aligned}$$

It means that a (real) number x satisfies the given equation if and only if $x = -5$ or $x = 3$. The statement "$x = -5$ or $x = 3$" cannot be replaced by "$x = -5$ and $x = 3$".

To say that there are two solutions, you may write "*the solutions are* -5 *and* 3". Sometimes, we also write "*the solutions are* $x_1 = -5$ *and* $x_2 = 3$" which means "*there are two solutions* -5 *and* 3 *and they are denoted by* x_1 *and* x_2 *respectively*".

In Chapter 1, you will learn the concept of *sets*. To specify a set, we may use "*listing*" or "*description*". The *solution set* to an equation is the set consisting of all the solutions to the equation. For the above example, we may write

- the solution set is $\{-5, 3\}$ *(listing)*;
- the solution set is $\{x : x = -5 \text{ or } x = 3\}$ *(description)*.

When we use *and*, we mean the listing method. □

Quadratic Formula Solutions to Equation (0.4.1) are given by

$$x = \frac{-b \pm \sqrt{b^2 - 4ac}}{2a}.$$

Remark $b^2 - 4ac$ is called the *discriminant* of (0.4.1).

(1) If $b^2 - 4ac > 0$, then (0.4.1) has two distinct solutions.

(2) If $b^2 - 4ac = 0$, then (0.4.1) has one solution.

(3) If $b^2 - 4ac < 0$, then (0.4.1) has no (real) solution.

FAQ Why is "(real)" added?

Answer When the real number system is enlarged to the complex number system, (0.4.1) has two complex solutions if $b^2 - 4ac < 0$. However, these solutions are not real numbers. In this course, we consider real numbers only. So you may simply say that there is no solution. □

Example Solve the following quadratic equations.

(1) $2x^2 - 9x + 10 = 0$

(2) $x^2 + 2x + 3 = 0$

Solution

(1) Using the quadratic formula, we see that the equation has two solutions given by

$$x = \frac{9 \pm \sqrt{(-9)^2 - 4(2)(10)}}{2(2)} = \frac{9 \pm 1}{4}.$$

Thus the solutions are $\frac{5}{2}$ and 2.

(2) Since $2^2 - 4(1)(3) = -8 < 0$, the equation has no solutions.

□

Example Solve the equation $x(x+2) = x(2x+3)$.

Solution Expanding both sides, we get

$$\begin{aligned} x^2 + 2x &= 2x^2 + 3x \\ x^2 + x &= 0 \\ x(x+1) &= 0 \\ x = 0 \quad &\text{or} \quad x = -1 \end{aligned}$$

The solutions are -1 and 0. □

Remark If we cancel the factor x on both sides, we get $x + 2 = 2x + 3$ which has only one solution. In canceling the factor x, it is assumed that $x \neq 0$. However, 0 is a solution and so this solution is lost. To use cancellation, we should write

$$x(x+2) = x(2x+3) \iff x + 2 = 2x + 3 \quad \text{or} \quad x = 0$$
$$\vdots$$

Example Find the value(s) of k such that the equation $3x^2 + kx + 7 = 0$ has only one solution.

Solution The given equation has only one solution iff

$$k^2 - 4(3)(7) = 0.$$

Solving, we get $k = \pm\sqrt{84}$. □

Exercise 0.4

1. Solve the following equations.

   (a) $4x - 4x^2 = 0$ (b) $2 + x - 3x^2 = 0$

   (c) $4x(x-4) = x - 15$ (d) $x^2 + 2\sqrt{2}x + 2 = 0$

   (e) $x^2 + 2\sqrt{2}x + 3 = 0$ (f) $x^3 - 7x^2 + 3x = 0$

2. Find the value(s) of k such that the equation $x^2 + kx + (k+3) = 0$ has only one solution.
3. Find the positive number such that sum of the number and its square is 210.

0.5 Remainder Theorem and Factor Theorem

Remainder Theorem *If a polynomial $p(x)$ is divided by $x - c$, where c is a constant, the remainder is $p(c)$.*

Example Let $p(x) = x^3 + 3x^2 - 2x + 2$. Find the remainder when $p(x)$ is divided by $x - 2$.

Solution The remainder is $p(2) = 2^3 + 3(2^2) - 2(2) + 2 = 18$. □

Factor Theorem $(x - c)$ *is a factor of a polynomial $p(x)$ if and only if $p(c) = 0$.*

Proof This follows immediately from the remainder theorem because $(x-c)$ is a factor means that the remainder is 0. □

Example Let $p(x) = x^3 + kx^2 + x - 6$. Suppose that $(x + 2)$ is a factor of $p(x)$.

(1) Find the value of k.

(2) With the value of k found in (1), factorize $p(x)$.

Solution

(1) Since $(x - (-2))$ is a factor of $p(x)$, it follows from the Factor Theorem that $p(-2) = 0$, that is

$$(-2)^3 + k(-2)^2 + (-2) - 6 = 0.$$

Solving, we get $k = 4$.

(2) Using long division, we get

$$x^3 + 4x^2 + x - 6 = (x + 2)(x^2 + 2x - 3).$$

By inspection, we have $p(x) = (x + 2)(x + 3)(x - 1)$. □

FAQ Can we find the quotient $(x^2 + 2x - 3)$ by inspection (without using long division)?

Answer The "inspection method" that some students use is called the *compare coefficient method*. Since the quotient is quadratic, it is in the form $(ax^2 + bx + c)$. Thus we have

$$x^3 + 4x^2 + x - 6 = (x + 2)(ax^2 + bx + c) \tag{0.5.1}$$

Comparing the coefficient of x^3, we see that $a = 1$. Similarly, comparing the constant term, we get $c = -3$. Hence we have

$$x^3 + 4x^2 + x - 6 = (x + 2)(x^2 + bx - 3).$$

To find b, we may compare the x^2 term (or the x term) to get

$$4 = 2 + b,$$

which yields $b = 2$.

Remark The compare coefficient method in fact consists of the following steps:

(1) Expand the right side of (0.5.1) to get

$$ax^3 + (2a + b)x^2 + (2b + c)x + 2c$$

(2) Compare the coefficients of the given polynomial with that obtained in Step (1) to get

$$\begin{aligned} 1 &= a \\ 4 &= 2a + b \\ 1 &= 2b + c \\ -6 &= 2c \end{aligned}$$

(3) Solve the above system to find a, b and c.

□

Example Factorize $p(x) = 2x^2 - 3x - 1$.

Solution Solving $p(x) = 0$ by the quadratic formula, we get

$$x = \frac{-(-3) \pm \sqrt{(-3)^2 - 4(2)(-1)}}{2(2)} = \frac{3 \pm \sqrt{17}}{4}.$$

By the Factor Theorem, both $\left(x - \frac{3+\sqrt{17}}{4}\right)$ and $\left(x - \frac{3-\sqrt{17}}{4}\right)$ are factors of $p(x)$. Therefore, we have

$$p(x) = 2\left(x - \frac{3+\sqrt{17}}{4}\right)\left(x - \frac{3-\sqrt{17}}{4}\right)$$

where the factor 2 is obtained by comparing the leading term (that is, the x^2 term). □

FAQ Can we say that $p(x)$ can't be factorized?

Answer Although $p(x)$ does not have factors in the form $(x - c)$ where c is an integer, it has linear factors as given above. If the question asks for factors with integer coefficients, then $p(x)$ cannot be factorized as product of linear factors. □

FAQ Can we use the above method to factorize, for example, $p(x) = 6x^2 + x - 2$?

Answer If you don't know how to factorize $p(x)$ by inspection, you can solve $p(x) = 0$ using the quadratic formula (or calculators) to get $x = \frac{1}{2}$ or $x = -\frac{2}{3}$. Therefore (by the Factor Theorem and comparing the leading term), we have

$$\begin{aligned} p(x) &= 6\left(x - \frac{1}{2}\right)\left(x + \frac{2}{3}\right) \\ &= (2x-1)(3x+2). \end{aligned}$$

□

Exercise 0.5

1. For each of the following expressions, use the factor theorem to find a linear factor $(x - c)$ and hence factorize it completely (using integer coefficients).

   (a) $x^3 - 13x + 12$ (b) $2x^3 - 7x^2 + 2x + 3$

   (c) $2x^3 - x^2 - 4x + 3$ (d) $x^3 - 5x^2 + 11x - 7$

2. Solve the following equation for x.

   (a) $2x^3 - 9x^2 - 8x + 15 = 0$ (b) $x^3 - 2x + 1 = 0$

   (c) $2x^3 - 5x^2 + 2x - 15 = 0$

0.6 Solving Linear Inequalities

Notation and Terminology Let a and b be real numbers.

(1) We say that b *is greater than* a, or equivalently, that a *is less than* b to mean that $b - a$ is a positive number.

(2) We write $b > a$ to denote that b is greater than a and we write $a < b$ to denote that a is less than b.

(3) We write $b \geq a$ to denote that b *is greater than or equal to* a and we write $a \leq b$ to denote that a *is less than or equal to* b.

A *linear inequality* in one unknown x is an inequality that can be written in one of the following forms:

(1) $ax + b < 0$

(2) $ax + b \leq 0$

(3) $ax + b > 0$

(4) $ax + b \geq 0$

where a and b are constants with $a \neq 0$. More generally, an inequality in one unknown x is an inequality that can be written in one of the following forms:

(1) $F(x) < 0$

(2) $F(x) \leq 0$

(3) $F(x) > 0$

(4) $F(x) \geq 0$

where F is a function from a subset of $\mathbb{R}$ into $\mathbb{R}$.

Definition A *solution* to an inequality $F(x) < 0$ is a real number x_0 such that $F(x_0) < 0$. The definition also applies to other types of inequalities.

Example Consider the inequality $2x + 3 \geq 0$. By direct substitution, we see that 1 is a solution and -2 is not a solution.

To solve an inequality means to find all solutions to the inequality.

Rules for Inequalities Let a, b and c be real numbers. Then the following holds.

(1) If $a < b$, then $a + c < b + c$.

(2) If $a < b$ and $c > 0$, then $ac < bc$.

(3) If $a < b$ and $c < 0$, then $ac > bc$. Note: The inequality is *reversed*.

(4) If $a < b$ and $b \leq c$, then $a < c$.

(5) If $a < b$ and a and b have the same sign, then $\frac{1}{a} > \frac{1}{b}$.

(6) If $0 < a < b$ and n is a positive integer, then $a^n < b^n$ and $\sqrt[n]{a} < \sqrt[n]{b}$.

Terminology Two numbers have the *same sign* means that both of them are positive or both of them are negative.

Remark One common mistake in solving inequalities is to apply a rule with the wrong sign (positive or negative). For example, if c is negative, it would be wrong to apply Rule (2).

Example Solve the following inequalities.

(1) $2x + 1 > 7(x + 3)$

(2) $3(x-2)+5 > 3x+7$

Solution

(1) Using rules for inequalities, we get

$$\begin{aligned} 2x+1 &> 7(x+3) \\ 2x+1 &> 7x+21 \\ 1-21 &> 7x-2x \\ -20 &> 5x \\ -4 &> x. \end{aligned}$$

The solutions are all the real numbers x such that $x < -4$, that is, all real numbers less than -4.

(2) Expanding the left side, we get

$$3(x-2)+5 = 3x-1$$

which is always less than the right side. Thus the inequality has no solution.

□

Exercise 0.6

1. Solve the following inequalities for x.

(a) $\frac{1-x}{2} \geq \frac{3x-7}{3}$ (b) $2(3-x) \leq \sqrt{3}(1-x)$

(c) $\frac{3x}{1-x} + 3 < 0$ (d) $\frac{2x}{2x+3} > 1$

0.7 Lines

A *linear equation in two unknowns* x and y is an equation that can be written in the form

$$ax + by + c = 0 \tag{0.7.1}$$

where a, b and c are constants with a, b not both 0. More generally, an *equation in two unknowns* x and y is an equation that can be written in the form

$$F(x, y) = 0, \tag{0.7.2}$$

where F is a function (from a collection of ordered pairs into $\mathbb{R}$).

Definition An *ordered pair (of real numbers)* is a pair of real numbers x_0, y_0 enclosed inside parenthesis: (x_0, y_0).

Remark Two ordered pairs (x_0, y_0) and (x_1, y_1) are equal if and only if $x_0 = x_1$ and $y_0 = y_1$. For example, the ordered pairs $(1, 2)$ and $(2, 1)$ are not equal.

Definition A *solution* to Equation (0.7.2) is an ordered pair (x_0, y_0) such that $F(x_0, y_0) = 0$.

Example Consider the equation

$$2x + 3y - 4 = 0.$$

By direct substitution, we see that $(2, 0)$ is a solution whereas $(1, 2)$ is not a solution.

Rectangular Coordinate System Given a plane, there is a one-to-one correspondence between points in the plane and ordered pairs of real numbers (see the construction below). The plane described in this way is called the *Cartesian plane* or the *rectangular coordinate plane*.

First we construct a horizontal line and a vertical line on the plane. Their point of intersection is called the *origin*. The horizontal line is called the x-axis and the vertical line y-axis. For each point P in the plane we can label it by two real numbers. To this ends, we draw perpendiculars from P to the x-axis and y-axis. The first perpendicular meets the x-axis at a point which can be represented by a real number a. Similarly, the second perpendicular meets the y-axis at a point which can be represented by a real number b. Moreover, the ordered pair of numbers a and b determines P uniquely, that is, if P_1 and P_2 are distinct points in the plane, then the ordered pairs corresponding to P_1 and P_2 are different. Therefore, we may identify the point P with the ordered pair (a, b) and we write $P = (a, b)$ or $P(a, b)$. The numbers a and b are called the x-coordinate and y-coordinate of P respectively.

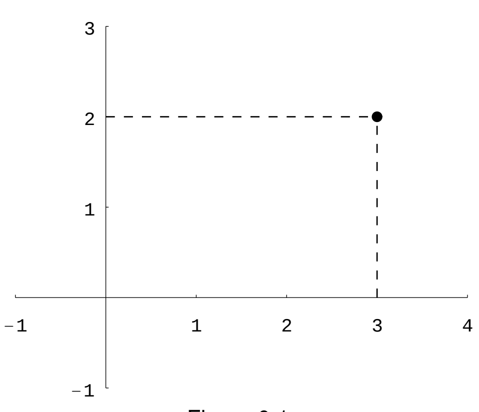


Figure 0.1

The x- and y-axes divide the (rectangular) coordinate plane into 4 regions (called *quadrants*):

$$\text{Quadrant I} = \{(a, b) : a > 0 \text{ and } b > 0\}, \qquad \text{Quadrant II} = \{(a, b) : a < 0 \text{ and } b > 0\},$$
$$\text{Quadrant III} = \{(a, b) : a < 0 \text{ and } b < 0\}, \qquad \text{Quadrant IV} = \{(a, b) : a > 0 \text{ and } b < 0\}.$$

Lines in the Coordinate Plane Consider the following equation

$$Ax + By + C = 0 \tag{0.7.3}$$

where A, B and C are constants with A, B not both zero. It is not difficult to see that the equation has infinitely many solutions. Each solution (x_0, y_0) represents a point in the (rectangular) coordinate plane. The collection of all solutions (points) form a line, called the *graph* of Equation (0.7.3). Moreover, every line in the plane can be represented in this way. For example, if ℓ is the line passing through the origin and making an angle of 45 degrees with the positive x-axis, then it is the graph of the equation $y = x$. Although this equation is not in the form (0.7.3), it can be written as

$$(1)x + (-1)y + 0 = 0,$$

that is, $x - y = 0$.

Terminology If a line ℓ is represented by an equation in the form (0.7.3), we say that the equation is a *general linear form* for ℓ.

Remark In Equation (0.7.3),

(1) if $A = 0$, then the equation reduces to $y = -\dfrac{C}{B}$ and its graph is a horizontal line;

(2) if $B = 0$, then the equation reduces to $x = -\dfrac{C}{A}$ and its graph is a vertical line.

Example Consider the line ℓ given by

$$2x + 3y - 4 = 0 \tag{0.7.4}$$

For each of the following points, determine whether it lies on ℓ or not.

(1) $A = (4, -1)$

(2) $B = (5, -2)$

Solution

(1) Putting $(x, y) = (4, -1)$ into (0.7.4), we get

$$\begin{aligned} L.S. &= 2(4) + 3(-1) - 4 = 1 \\ &\neq 0. \end{aligned}$$

Therefore A does not lie on ℓ.

(2) Putting $(x, y) = (5, -2)$ into (0.7.4), we get

$$\begin{aligned} L.S. &= 2(5) + 3(-2) - 4 = 0 \\ &= R.S. \end{aligned}$$

Therefore B lies on ℓ.

□

Example Consider the line ℓ given by

$$x + 2y - 4 = 0 \tag{0.7.5}$$

Find the points of intersection of ℓ with the x-axis and the y-axis.

Solution

- Putting $y = 0$ into (0.7.5), we get

$$x - 4 = 0$$

  from which we obtain $x = 4$.

  The point of intersection of ℓ with the x-axis is $(4, 0)$.

- Putting $x = 0$ into (0.7.5), we get

$$2y - 4 = 0$$

  from which we obtain $y = 2$.

  The point of intersection of ℓ with the y-axis is $(0, 2)$.

□

Remark The point $(4, 0)$ and $(0, 2)$ are called the x-intercept and y-intercept of ℓ respectively.

FAQ Can we say that the x-intercept is 4 etc?

Answer Some authors define x-intercept to be the x-coordinate of point of intersection etc. Using this convention, the x-intercept is 4 and the y-intercept is 2. □

Definition For a non-vertical line ℓ, its slope (denoted by m_ℓ or simply m) is defined to be

$$m_\ell = \frac{y_2 - y_1}{x_2 - x_1}$$

where $P_1(x_1, y_1)$ and $P_2(x_2, y_2)$ are any two distinct points lying on ℓ.

Remark The number m_ℓ is well-defined, that is, its value is independent of the choice of P_1 and P_2.

FAQ What is the slope of a vertical line?

Answer The slope of a vertical line is undefined because if $P_1 = (x_1, y_1)$ and $P_2 = (x_2, y_2)$ lie on a vertical line, then $x_1 = x_2$ and so

$$\frac{y_2 - y_1}{x_2 - x_1} = \frac{y_2 - y_1}{0}$$

which is undefined.

Some students say that the slope is *infinity*, denoted by ∞. However, ∞ is not a number; it is just a notation. Moreover, infinity is ambiguous—does it mean positive infinity (going up, very steep) or negative infinity (going down, very steep)? □

Example Find the slope of the line (given by) $2x - 5y + 9 = 0$.

Solution Take any two points on the line, for example, take $P_1 = (-2, 1)$ and $P_2 = (3, 3)$. The slope m of the line is

$$m = \frac{3 - 1}{3 - (-2)} = \frac{2}{5}.$$

FAQ Can we take other points on the line?

Answer You can take any two points. For example, taking $A = (0, \frac{9}{5})$ and $B = (-\frac{9}{2}, 0)$, we get

$$m = \frac{\frac{9}{5} - 0}{0 - (-\frac{9}{2})} = \frac{\frac{9}{5}}{\frac{9}{2}} = \frac{2}{5}.$$

□

Equations for Lines

Let ℓ be a non-vertical line in the coordinate plane.

- Suppose $P = (x_1, y_1)$ is a point lying on ℓ and m is the slope of ℓ. Then an equation for ℓ can be written in the form

  $$y - y_1 = m(x - x_1) \tag{0.7.6}$$

  called a *point-slope form* for ℓ.

  Remark Since there are infinitely many points on a line, ℓ has infinitely many point-slope forms. However, we also say that (0.7.6) is *the* point-slope form of ℓ.

  FAQ Can we write the equation in the following form?

  $$\frac{y - y_1}{x - x_1} = m \tag{0.7.7}$$

  Answer Equation (0.7.7) represents a line minus one point. If you put $(x, y) = (x_1, y_1)$ into (0.7.7), the left-side is $\frac{0}{0}$ which is undefined. This means that the point (x_1, y_1) does not lie on L. However, once you get (0.7.7), you can obtain the point-slope form (0.7.6) easily. □

- Suppose the y-intercept of ℓ is $(0, b)$ and the slope of ℓ is m. Then a point-slope form for ℓ is

$$y - b = m(x - 0)$$

which can be written as

$$y = mx + b$$

called the *slope-intercept form* for ℓ.

Example Find the slope of the line having general linear form $2x + 3y - 4 = 0$.

Solution Rewrite the given equation in slope-intercept form:

$$\begin{aligned} 2x + 3y - 4 &= 0 \\ 3y &= -2x + 4 \\ y &= -\frac{2}{3}x + \frac{4}{3} \end{aligned}$$

The slope of the line is $-\frac{2}{3}$. □

Example Let ℓ be the line that passes through the points $A(1, 3)$ and $B(2, -4)$. Find an equation in general linear form for ℓ.

Solution Using the points A and B, we get the slope m of ℓ

$$m = \frac{3 - (-4)}{1 - 2} = -7.$$

Using the slope m and the point A (or B), we get the point slope form

$$y - 3 = -7(x - 1). \tag{0.7.8}$$

Expanding and rearranging terms, (0.7.8) can be written in the following general linear form

$$7x + y - 10 = 0.$$

□

Parallel and Perpendicular Lines Let ℓ_1 and ℓ_2 be (non-vertical) lines with slopes m_1 and m_2 respectively. Then

(1) ℓ_1 and ℓ_2 are parallel if and only if $m_1 = m_2$;

(2) ℓ_1 and ℓ_2 are perpendicular to each other if and only if $m_1 \cdot m_2 = -1$.

Note

- If ℓ_1 and ℓ_2 are vertical, then they are parallel.
- If ℓ_1 is vertical and ℓ_2 is horizontal (or the other way round), then they are perpendicular to each other.

Example Find equations in general linear form for the two lines passing through the point $(3, -2)$ such that one is parallel to the line $y = 3x + 1$ and the other is perpendicular to it.

Solution Let ℓ_1 (respectively ℓ_2) be the line that passes through the point $(3, -2)$ and parallel (respectively perpendicular) to the given line. It is clear that the slope of the given line is 3. Thus the slope of ℓ_1 is 3 and the slope of ℓ_2 is $-\frac{1}{3}$. From these, we get the point-slope forms for ℓ_1 and ℓ_2:

$$y - (-2) = 3(x - 3) \qquad \text{and} \qquad y - (-2) = -\frac{1}{3}(x - 3)$$

respectively. Expanding and rearranging terms, we get the following linear forms

$$3x - y - 11 = 0 \qquad \text{and} \qquad x + 3y + 3 = 0$$

for ℓ_1 and ℓ_2 respectively. □

Exercise 0.7

1. For each of the following, find an equation of the line satisfying the given conditions. Give your answer in general linear form.
   (a) Passing through the origin and $(-2, 3)$.
   (b) With slope 2 and passing through $(5, -1)$.
   (c) With slope -3 and y-intercept $(0, 7)$.
   (d) Passing through $(-3, 2)$ and parallel to $2x - y - 3 = 0$.
   (e) Passing through $(1, 4)$ and perpendicular to $x + 3y = 0$.
   (f) Passing through $(1, -1)$ and perpendicular to the y-axis.

0.8 Pythagoras Theorem, Distance Formula and Circles

Pythagoras Theorem *Let a, b and c be the (lengths of the) sides of a right-angled triangle where c is the hypotenuse. Then we have*

$$a^2 + b^2 = c^2.$$

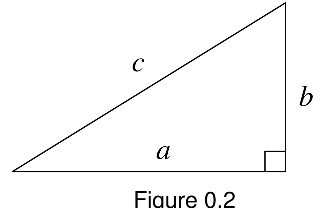


Figure 0.2

Distance Formula Let $P = (x_1, y_1)$ and $Q = (x_2, y_2)$. Then the distance PQ between P and Q is

$$PQ = \sqrt{(x_2 - x_1)^2 + (y_2 - y_1)^2}.$$

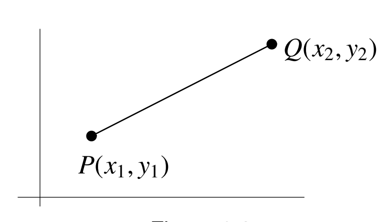


Figure 0.3

Equation of Circles Let $\mathcal{C}$ be the circle with center at $C(h,k)$ and radius r. Then an equation for $\mathcal{C}$ is

$$(x-h)^2+(y-k)^2=r^2. \tag{0.8.1}$$

Proof Let $P(x,y)$ be any point on the circle. Since the distance from P to the center C is r, using the distance formula, we get

$$\sqrt{(x-h)^2+(y-k)^2}=r.$$

Squaring both sides yields (0.8.1). □

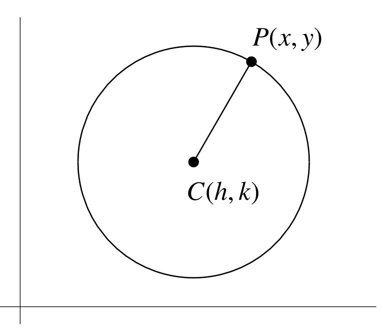


Figure 0.4

Example Find the center and radius of the circle given by

$$x^2-4x+y^2+6y-12=0.$$

Solution Using the completing square method, the given equation can be written in the form (0.8.1).

$$\begin{aligned} x^2-4x+y^2+6y &= 12 \\ (x^2-4x+4)+(y^2+6y+9) &= 12+4+9 \\ (x-2)^2+(y+3)^2 &= 25 \\ (x-2)^2+(y-(-3))^2 &= 5^2 \end{aligned}$$

The center is $(2,-3)$ and the radius is 5. □

FAQ How do we get the number "9" etc (the numbers added to both sides)?

Answer We want to find a number (denoted by a) such that (y^2+6y+a) is a complete square. That is,

$$y^2+6y+a=(y+b)^2 \tag{0.8.2}$$

for some number b. Expanding the right-side of (0.8.2) (do this in your head) and comparing the coefficients of y on both sides, we get $2b=6$, that is, $b=3$. Hence comparing the constant terms on both sides, we get $a=b^2=9$.

Summary a = square of half of the coefficient of y. □

Exercise 0.8

1. For each of the following pairs of points, find the distance between them.
   (a) $(-3,4)$ and the origin (b) $(4,0)$ and $(0,-7)$
   (c) $(7,5)$ and $(12,17)$ (d) $(-2,9)$ and $(3,-1)$
2. For each of the following circles, find its radius and center.
   (a) $x^2+y^2-4y+1=0$ (b) $x^2+y^2+4x-2y-4=0$
   (c) $2x^2+2y^2+4x-2y+1=0$
3. For each of the following, find the distance from the given point to the given line.
   (a) $(-2,3)$ and the y-axis
   (b) the origin and $x+y=1$
   (c) $(1,2)$ and $2x+y-6=0$

0.9 Parabola

The graph of

$$y = ax^2 + bx + c$$

where $a \neq 0$, is a *parabola*. The parabola intersects the x-axis at two distinct points if $b^2 - 4ac > 0$. It touches the x-axis (one intersection point only) if $b^2 - 4ac = 0$ and does not intersect the x-axis if $b^2 - 4ac < 0$.

- If $a > 0$, the parabola opens upward and there is a lowest point (called the *vertex* of the parabola).
- If $a < 0$, the parabola opens downward and there is a highest point (*vertex*).

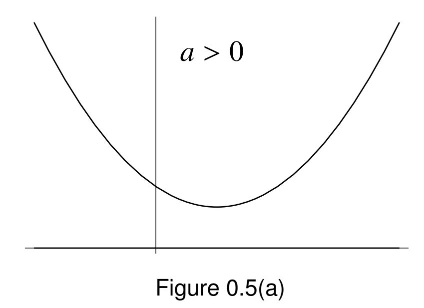


Figure 0.5(a)

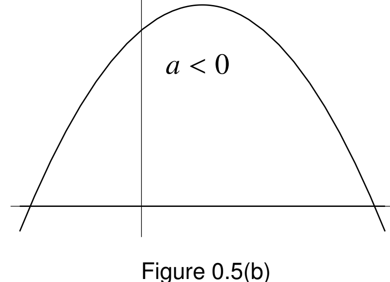


Figure 0.5(b)

The vertical line that passes through the vertex is called the *axis of symmetry* because the parabola is symmetric about this line.

To find the vertex, we can use the completing square method to write the equation in the form

$$y = a(x - h)^2 + k \tag{0.9.1}$$

The vertex is (h, k) because $(x - h)^2$ is always non-negative and so

- if $a > 0$, then $y \geq k$ and thus (h, k) is the lowest point;
- if $a < 0$, then $y \leq k$ and thus (h, k) is the highest point.

Example Consider the parabola given by

$$y = x^2 + 6x + 5.$$

Find its vertex and axis of symmetry.

Solution Using the completing square method, the given equation can be written in the form (0.9.1).

$$\begin{aligned} y &= x^2 + 6x + 5 \\ y &= (x^2 + 6x + 9) - 9 + 5 \\ y &= (x + 3)^2 - 4 \\ y &= (x - (-3))^2 - 4. \end{aligned}$$

The vertex is $(-3, -4)$ and the axis of symmetry is the line given by $x = -3$ (the vertical line that passes through the vertex). □

FAQ In the above example, the coefficient of x^2 is 1, what should we do if it is not 1?

Answer To illustrate the procedure, let's consider $y = 2x^2 + 3x - 4$. To rewrite the equation in the form (0.9.1), consider the first two terms and rewrite it in the form $a(x^2 + \frac{b}{a}x)$.

$$\begin{aligned}
y &= 2\left(x^2 + \frac{3}{2}x\right) - 4 \\
y &= 2\left(x^2 + \frac{3}{2}x + \left(\frac{3}{4}\right)^2 - \left(\frac{3}{4}\right)^2\right) - 4 \\
y &= 2\left(x^2 + \frac{3}{2}x + \left(\frac{3}{4}\right)^2 - \frac{9}{16}\right) - 4 \\
y &= 2\left(x + \frac{3}{4}\right)^2 - \frac{9}{8} - 4 \\
y &= 2\left(x + \frac{3}{4}\right)^2 - \frac{41}{8}
\end{aligned}$$

□

Exercise 0.9

1. For each of the following parabolas, find its x-intercept(s), y-intercept and vertex.

   (a) $y = x^2 + 4x - 12$ (b) $y = -x^2 + 6x - 7$

   (c) $y = 2x^2 + 2x + 7$

0.10 Systems of Equations

A system of two equations in two unknowns x and y can be written as

$$\begin{aligned}
F_1(x, y) &= 0 \\
F_2(x, y) &= 0.
\end{aligned}$$

Usually, each equation represents a curve in the coordinate plane. Solving the system means to find all ordered pairs (x_0, y_0) such that $F_1(x_0, y_0) = 0$ and $F_2(x_0, y_0) = 0$, that is, to find all points $P(x_0, y_0)$ that lies on the intersection of the two curves.

To solve a system of two linear equations (with two unknowns x and y)

$$\begin{aligned}
ax + by + c &= 0 \\
dx + ey + f &= 0,
\end{aligned}$$

we can use elimination or substitution.

Example Solve the following system of equations

$$\begin{aligned}
2x + 3y &= 7 \qquad (0.10.1) \\
3x + 5y &= 11 \qquad (0.10.2)
\end{aligned}$$

Solution

(*Elimination*) Multiply (0.10.1) and (0.10.2) by 3 and 2 respectively, we get

$$\begin{aligned}
6x + 9y &= 21 \qquad (0.10.3) \\
6x + 10y &= 22 \qquad (0.10.4)
\end{aligned}$$

Subtracting (0.10.3) from (0.10.4), we get $y = 1$.

Substituting $y = 1$ back into (0.10.1) or (0.10.2) and solving, we get $x = 2$.

The solution to the system is $(2, 1)$.

Remark The point $(2, 1)$ is the intersection point of the lines given by $2x + 3y = 7$ and $3x + 5y = 11$.

(*Substitution*) From (0.10.1), we get $x = \frac{7-3y}{2}$. Substituting into (0.10.2), we get

$$\begin{aligned} 3\left(\frac{7-3y}{2}\right) + 5y &= 11 \\ 3(7-3y) + 10y &= 22 \\ y &= 1 \end{aligned}$$

and we can proceed as in the elimination method.

□

To solve a system in two unknowns, with one linear equation and one quadratic equation

$$\begin{aligned} ax + by + c &= 0 \\ dx^2 + exy + fy^2 + gx + hy + k &= 0 \end{aligned}$$

we can use substitution. From the linear equation, we can express x in terms of y (or vice versa). Substituting into the quadratic equation, we get a quadratic equation in y which can be solved by factorization or by formula. Substituting the value(s) of y back into the linear equation, we get the corresponding value(s) of x.

Example Solve the following system of equations

$$x - 2y = 4 \tag{0.10.5}$$

$$x^2 + y^2 = 5 \tag{0.10.6}$$

Solution From (0.10.5), we get $x = 4 + 2y$. Substituting into (0.10.6), we get

$$\begin{aligned} (4+2y)^2 + y^2 &= 5 \\ 5y^2 + 16y + 11 &= 0. \end{aligned}$$

Solving we get $y = -1$ or $y = -\frac{11}{5}$.

Substituting $y = -1$ into (0.10.5), we get $x = 2$; substituting $y = -\frac{11}{5}$ into (0.10.5), we get $x = -\frac{2}{5}$.

The solutions to the system are $(2, -1)$ and $(-\frac{2}{5}, -\frac{11}{5})$. □

Remark

- If we substitute $y = -1$ into (0.10.6), we get two values of x, one of which should be rejected.
- The solutions are the intersection points of the line $x - 2y = 4$ and the circle $x^2 + y^2 = 5$.

Example Find the point(s) of intersection, if any, of the line and the parabola given by $x + y - 1 = 0$ and $y = x^2 + 2$ respectively.

Solution From the equation of the line, we get $y = 1 - x$. Putting into the equation of the parabola, we get

$$\begin{aligned} 1 - x &= x^2 + 2 \\ 0 &= x^2 + x + 1. \end{aligned}$$

Since $\Delta = 1^2 - 4(1)(1) < 0$, the above quadratic equation has no solution. Hence the system

$$\begin{aligned} x + y - 1 &= 0 \\ y &= x^2 + 2 \end{aligned}$$

has no solution, that is, the line and the parabola do not intersect. □

Exercise 0.10

1. Consider a rectangle with perimeter 28 *cm* and diagonal 10 *cm*. Find the length and width of the rectangle.

Chapter 1

Sets, Real Numbers and Inequalities

1.1 Sets

1.1.1 Introduction

Idea of definition A *set* is a collection of objects.

This is not a definition because we have not defined what a *collection* is. If we give a definition for *collection*, it must involve something that have not been defined. It is impossible to define everything. In mathematics, *set* is a fundamental concept that cannot be defined. The idea of definition given above describes what a set is using daily language. This helps us "*understand*" the meaning of a set.

Terminology An object in a set is called an *element* or a *member* of the set.

To describe sets, we can use *listing* or *description*.

[*Listing*] To denote a set with finitely many elements, we can list all the elements of the set and enclose them by braces. For example,

$$\{1, 2, 3\}$$

is the set which has exactly three elements, namely 1, 2 and 3.

If we want to denote the set whose elements are the first one hundred positive integers, it is impractical to write down all the elements. Instead, we write

$$\{1, 2, 3, \dots, 99, 100\}, \quad \text{or simply} \quad \{1, 2, \dots, 100\}.$$

The three dots "$\dots$"(read "*and so on*") means that the pattern is repeated, up to the number(s) listed at the end.

Suppose in a problem, we consider a set, say $\{1, 2, \dots, 100\}$. We may have to refer to the set later many times. Instead of writing $\{1, 2, \dots, 100\}$ repeatedly, we can give it a name by using a symbol to represent the set. Usually, we use small letters (eg. $a, b, \dots$) to denote objects and capital letters (eg. $A, B, \dots$) to denote sets. For example, we may write

- "Let $A = \{1, 2, \dots, 100\}$."

which means that the set $\{1, 2, \ldots, 100\}$ is given the "name" A. If we want to refer to the set later, we can just write A. For example,

- "Let $A = \{1, 2, \ldots, 100\}$. Then 100 is an element of A, but 101 is not an element of A."

If we consider another set, say $\{1, 2, 3, 4, 5\}$ and want to give it a name, we must not use the symbol A again, because in the problem, A always means the set $\{1, 2, \ldots, 100\}$. For example,

- "Let $A = \{1, 2, \ldots, 100\}$. Let $B = \{1, 2, 3, 4, 5\}$. Then every element of B is also an element of A. But there are elements of A that are not elements of B."

Remark The equality sign "=" can be used in several ways as the following examples illustrate.

(1) $1 + 2 = 3$.

(2) $x^2 + 1 = 5$.

(3) Let $A = \{1, 2, 3\}$.

The equality sign in (1) means equality of two quantities: the quantity on the left and the quantity on the right are equal.

The equality sign in (2) is an equality in an equation. It is true when $x = 2$ (for example) and it is not true when $x = 1$ (for example). Instead of using the equality sign, some authors use "==". The equation in (2) may be written as

(2′) $x^2 + 1 == 5$.

The equality sign in (3) has a different meaning. The sentence in (3) means that the set $\{1, 2, 3\}$ is denoted by A. The symbol "=" assigns a name to an object *(a set is also an object)*. The name is written on the left side and the object on the right side. Instead of using the equality sign, some authors use the symbol ":=". The sentence in (3) may be written as

(3′) Let $A := \{1, 2, 3\}$.

In this course, we will not use the notations ":=" and "==". Readers can determine the meaning of "=" from the context.

Notation Given an object x and a set A, either x is an element of A or x is not an element of A.

(1) If x is an element of A, we write $x \in A$ (read "*x belongs to A*").

(2) If x is not an element of A, we write $x \notin A$ (read "*x does not belong to A*").

There is a set that has no element. It is called the *empty set*, denoted by $\emptyset$. This is a Scandinavian letter, a zero 0 together with a slash /.

Definition The set that has no element is called the *empty set* and is denoted by $\emptyset$.

Remark Because the empty set has no element, if we list all the elements of it and enclose "them" by braces, we get { }. This is an alternative notation for the empty set.

[*Description*] Another way to denote a set is to describe a common property of the elements of the set, using the following notation:

$$\{x : P(x)\} \qquad \text{or} \qquad \{x \mid P(x)\}$$

read "*the set of all x such that P(x) (is true)*". For example, the set whose elements are the first one hundred positive integers can be expressed as

(†) $\{x : x \text{ is a positive integer less than } 101\}$

In considering "property", it is understood that the property applies to a certain collection of objects only. For example, when we say "*an old person*" (a person is said to be old if his or her age is 65 or above), the property of being "*old*" is applied to people. It is meaningless to say "*this is an old atom*" (unless we have a definition which tells whether an atom is old or not).

The property of being a positive integer less than 101 is applied to numbers. In this course, we consider real numbers only. The set of all real numbers is denoted by $\mathbb{R}$. In considering the set given in (†), it is understood that x is a real number. To make this explicit, we write

(‡) $\{x \in \mathbb{R} : x \text{ is a positive integer less than } 101\}$

read "*the set of all x belonging to* $\mathbb{R}$ *such that x is a positive integer less than* 101".

Notation

(1) The set of all real numbers is denoted by $\mathbb{R}$.

(2) The set of all rational numbers is denoted by $\mathbb{Q}$.

(3) The set of all integers is denoted by $\mathbb{Z}$.

(4) The set of all positive integers is denoted by $\mathbb{Z}_+$.

(5) The set of all natural numbers is denoted by $\mathbb{N}$.

Definition

(1) A *rational number* is a number that can be written in the form $\frac{p}{q}$ where p and q are integers and $q \neq 0$.

(2) Positive integers together with 0 are called *natural numbers*.

Remark Some authors do not include 0 as natural number. In that case, $\mathbb{N}$ means the set of all positive integers.

Example

(1) To say that 2 is a natural number, we may write $2 \in \mathbb{N}$.

(2) To say that 2 is a rational number, we may write $2 \in \mathbb{Q}$.
Note: The number 2 is a rational number because it can be written as $\frac{2}{1}$ or $\frac{6}{3}$ etc.

(3) To say that π is not a rational number, we may write $\pi \notin \mathbb{Q}$.
Note: $\pi \neq \frac{22}{7}$; the rational number $\frac{22}{7}$ is only an approximation to π.

Definition Let A and B be sets. If every element of A is also an element of B and vice versa, then we say that A and B are *equal*, denoted by $A = B$.

Remark

- In mathematics, definitions are important. Students who want to take more courses in mathematics must pay attention to definitions. Understand the meaning, give examples, give nonexamples.

- In the definition, the first sentence "*Let A and B be sets*" describes the setting. The definition for equality applies to *sets* only and does not apply to other objects. Of course, we can consider equality of other objects, but it is another definition.
- In the first sentence "*Let A and B be sets*", the use of plural "*sets*" does not mean that A and B are two different sets. It also includes the case where A and B are the same set. The following are alternative ways to say this:
  - ◊ *Let A and B be set(s).*
  - ◊ *Let A be a set and let B be a set.*

  However, these alternative ways are rather cumbersome and will not be used in most situations.
- Some students may not be familiar with the use of the word "*let*". It is used very often in mathematics. Consider the following sentences:
  - ◊ Let $A = \{1, 2, 3, 4, 5\}$.
  - ◊ Let A be a set.

  The word "*let*" appears in both sentences. However, the meanings of "*let*" in the two sentences are quite different. In the first sentence, "*let*" means *denote* whereas in the second sentence, it means *suppose*. The definition for equality of sets can also be stated in the following ways:
  - ◊ Suppose A and B are sets. If every element of A is also an element of B and vice versa, then we say that A and B are *equal*.
  - ◊ If A and B are sets and if every element of A is also an element of B and vice versa, then we say that A and B are *equal*.
- The definition can also be stated in a way that the assumption that A and B are sets is combined with the condition for equality of A and B.
  - ◊ If every element of a set A is also an element of a set B and vice versa, then we say that A and B are *equal*.
- The definition tells that if A and B are sets having the same elements, then $A = B$. Conversely, it also tells that if A and B are sets and $A = B$, then A and B have the same elements because this is the condition to check whether A and B are equal. Some mathematicians give the definition using iff:
  - ◊ Let A and B be sets. We say that A and B are *equal* if and only if every element of A is also an element of B and vice versa.
- Sometimes, we also give definition of a concept together with its "*opposite*". The following is a definition of equality of sets together with its opposite. In this course, we will use the following format
  (a) *describe the setting*;
  (b) *give condition(s) for the concept*;
  (c) *give condition(s) for the opposite concept*,

  whenever it is appropriate.

Definition Let A and B be sets. If every element of A is also an element of B and vice versa, then we say that

A and B are *equal*, denoted by $A = B$. Otherwise, we say that A and B are *unequal*, denoted by $A \neq B$.

Example Let $A = \{1, 3, 5, 7, 9\}$ and let $B = \{x \in \mathbb{Z} : x \text{ is a positive odd number less than } 10\}$. Then we have $A = B$, that is, A and B are equal. This is because every element of A is also an element of B and vice versa.
Recall: $\mathbb{Z}$ is the set of all integers. Thus B is the set of all integers that are positive, odd and less than 10.

Remark To prove that the sets A and B in the above example are equal, we check whether the condition given in the definition is satisfied. This is called *proof by definition*.

Example Let $A = \{1, 3, 5, 7, 9\}$ and let $B = \{x \in \mathbb{Z}_+ : x \text{ is a prime number less than } 10\}$. Then we have $A \neq B$.
Recall: $\mathbb{Z}_+$ is the set of all positive integers.

Proof The number 9 is an element of A, but it is not an element of B. Therefore, it is not true that every element of A is also an element of B. Hence we have $A \neq B$. □

FAQ In the above two examples, the assertions are quite obvious. Do we need to prove them?

Answer Sometimes, mathematicians also write "*obvious*" in proofs of theorems. To some people, a result may be obvious; but, it may not be obvious to other people. If you say obvious, make sure that it is really obvious—if your classmates ask you why, you should be able to explain to them.

It is impractical to explain everything. In proving theorems or giving solutions to examples, reasons that are "*obvious*" will not be given. When you answer questions, you should use your own judgment. □

Remark Because it is impractical (in fact, impossible) to explain everything, discussion below will not be so detail as that above. If you don't understand a concept, read the definition again. Try different ways to understand it. Relate it with what you have learnt. Guess what the meaning is. See whether your guess is correct if you apply it to examples . . .

Example Let $A = \{1, 2, 3\}$ and let $B = \{1, 3, 2\}$. Then we have $A = B$.

Proof Obvious (use definition). □

The above example shows that in listing elements of a set, order is not important. It should also be noted that in listing elements, there is no need to repeat the elements. For example, $\{1, 2, 3, 2, 1\}$ and $\{1, 2, 3\}$ are the same set.

Definition Let A and B be sets. If every element of A is also an element of B, then we say that A is a *subset* of B, denoted by $A \subseteq B$. Otherwise, we say that A is *not a subset* of B, denoted by $A \nsubseteq B$.

Note

(1) $A \subseteq A$.

(2) $A = B$ if and only if $A \subseteq B$ and $B \subseteq A$.

(3) $A \nsubseteq B$ means that there is at least one element of A that is not an element of B.

Remark Instead of $A \subseteq B$, some authors use $A \subset B$ to denote A is a subset of B.

Example Let $A = \{1, 2, 3, 4, 5\}$, $B = \{1, 3, 5\}$ and $C = \{2, 4, 6\}$. Then we have $B \subseteq A$ and $C \nsubseteq A$.

The relation between A, B and C can be described by the diagram shown in Figure 1.1.

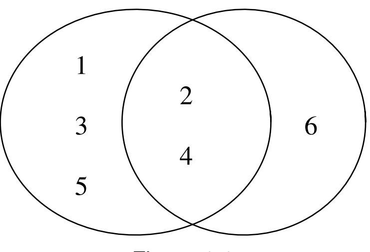


Figure 1.1

FAQ For the given sets A, B and C, we also have the following:

(1) $A \nsubseteq B$

(2) $A \nsubseteq C$

(3) $C \nsubseteq B$

(4) $B \nsubseteq C$

Why are they omitted?

Answer Good and correct observation. Given three sets, there are six ways to pair them up. The example just illustrates the meaning of $\subseteq$ and $\nsubseteq$. □

1.1.2 Set Operations

Definition Let A and B be sets.

(1) The *intersection* of A and B, denoted by $A \cap B$, is the set whose elements are those belonging to both A and B, that is,

$$A \cap B = \{x : x \in A \text{ and } x \in B\}.$$

(2) The *union* of A and B, denoted by $A \cup B$, is the set whose elements are those belonging to either A or B or both A and B, that is

$$A \cup B = \{x : x \in A \text{ or } x \in B\}.$$

Remark In mathematics, "P or Q" means "either P or Q or both P and Q".

Example Let $A = \{2, 3, 5\}$, $B = \{2, 5, 6, 8\}$ and $C = \{1, 2, 3\}$. Find the following sets.

(1) $A \cap B$

(2) $A \cup B$

(3) $(A \cap B) \cap C$

(4) $A \cap (B \cap C)$

Solution

(1) $A \cap B = \{2, 5\}$

(2) $A \cup B = \{2, 3, 5, 6, 8\}$

(3) $$\begin{aligned}(A \cap B) \cap C &= \{2, 5\} \cap \{1, 2, 3\} \\ &= \{2\}\end{aligned}$$

(4)
$$\begin{aligned} A \cap (B \cap C) &= \{2, 3, 5\} \cap \{2\} \\ &= \{2\} \end{aligned}$$
□

Note Given any sets A, B and C, we always have

$$(A \cap B) \cap C = A \cap (B \cap C) \qquad \text{and} \qquad (A \cup B) \cup C = A \cup (B \cup C).$$

Thus we may write $A \cap B \cap C$ and $A \cup B \cup C$ without ambiguity. We say that set intersection and set union are *associative*.

Definition Let A and B be sets. The *relative complement* of B in A, denoted by $A \setminus B$ or $A - B$ (read "*A setminus (or minus) B*"), is the set whose elements are those belonging to A but not belonging to B, that is,

$$A \setminus B = \{x \in A : x \notin B\}.$$

Example Let $A = \{a, b, c\}$ and $B = \{c, d, e\}$. Then we have $A \setminus B = \{a, b\}$.

For each problem, we will consider a set that is "large" enough, containing all objects under consideration. Such a set is called a *universal set* and is usually denoted by U. In this case, all sets under consideration are subsets of U and they can be written in the form $\{x \in U : P(x)\}$.

Example In considering addition and subtraction of whole numbers $(0, 1, 2, 3, 4, \ldots)$, we may use $\mathbb{Z}$ (the set of all integers) as a universal set.

(1) The set of all positive even numbers can be written as $\{x \in \mathbb{Z} : x > 0 \text{ and } x \text{ is divisible by } 2\}$.

(2) The set of all prime numbers can be written as $\{x \in \mathbb{Z} : x > 0 \text{ and } x \text{ has exactly two divisors}\}$.

Definition Let U be a universal set and let B be a subset of U. Then the set $U \setminus B$ is called the *complement* of B (in U) and is denoted by B' (or B^c).

Example Let $U = \mathbb{Z}_+$, the set of all positive integers. Let B be the set of all positive even numbers. Then B' is the set of all positive odd numbers.

Example Let $U = \{1, 2, 3, \ldots, 12\}$ and let

$$\begin{aligned} A &= \{x \in U : x \text{ is a prime number}\} \\ B &= \{x \in U : x \text{ is an even number}\} \\ C &= \{x \in U : x \text{ is divisible by } 3\}. \end{aligned}$$

Find the following sets.

(1) $A \cup B$

(2) $A \cap C$

(3) $B \cap C$

(4) $(A \cup B) \cap C$

(5) $(A \cap C) \cup (B \cap C)$

(6) $(A \cup B)'$

(7) $A' \cap B'$

Solution Note that

$$\begin{aligned} A &= \{2, 3, 5, 7, 11\} \\ B &= \{2, 4, 6, 8, 10, 12\} \\ C &= \{3, 6, 9, 12\}. \end{aligned}$$

(1) $A \cup B = \{2, 3, 4, 5, 6, 7, 8, 10, 11, 12\}$

(2) $A \cap C = \{3\}$

(3) $B \cap C = \{6, 12\}$

(4) $$\begin{aligned} (A \cup B) \cap C &= \{2, 3, 4, 5, 6, 7, 8, 10, 11, 12\} \cap \{3, 6, 9, 12\} \\ &= \{3, 6, 12\} \end{aligned}$$

(5) $$\begin{aligned} (A \cap C) \cup (B \cap C) &= \{3\} \cup \{6, 12\} \\ &= \{3, 6, 12\} \end{aligned}$$

(6) $$\begin{aligned} (A \cup B)' &= \{2, 3, 4, 5, 6, 7, 8, 10, 11, 12\}' \\ &= \{1, 9\} \end{aligned}$$

(7) $$\begin{aligned} A' \cap B' &= \{1, 4, 6, 8, 9, 10, 12\} \cap \{1, 3, 5, 7, 9, 11\} \\ &= \{1, 9\} \end{aligned}$$

□

Remark In the above example, we have

$$(A \cup B)' = A' \cap B' \qquad \text{and} \qquad (A \cup B) \cap C = (A \cap C) \cup (B \cap C).$$

In fact, these equalities are true in general.

Venn Diagrams

A *Venn diagram* is a very useful and simple device to represent sets graphically.

In a Venn diagram, the universal set U is usually represented by a rectangle. Inside this rectangle, subsets of the universal set are represented by circles, rectangles, or some other geometrical figures.

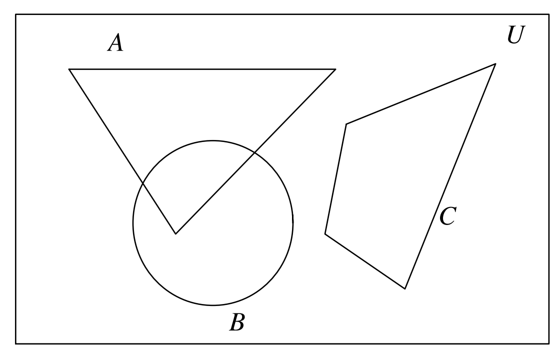


Figure 1.2

We can use Venn diagrams to obtain useful formulas for set operations.

- In Figure 1.3(a), the portion shaded by horizontal lines represents $A \cup B$ and that by vertical lines represents C; thus the portion shaded by both horizontal and vertical lines represents $(A \cup B) \cap C$.
- In Figure 1.3(b), the portion shaded by horizontal lines represents $A \cap C$ and that by vertical lines represents $B \cap C$; thus the portion shaded by vertical or horizontal lines represents $(A \cap C) \cup (B \cap C)$.

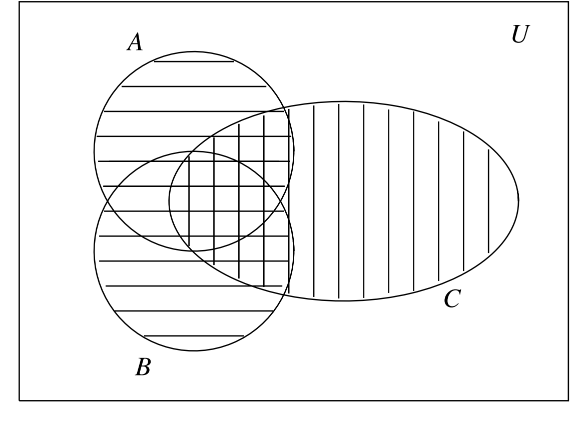


Figure 1.3(a)

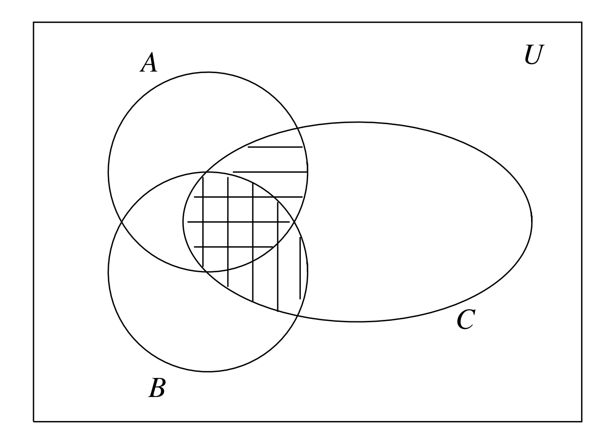


Figure 1.3(b)

From the two figures, we see that

$$(A \cup B) \cap C = (A \cap C) \cup (B \cap C).$$

Venn diagrams help us in a visual way to identify the above formulas. However, in order to prove these formulas in a rigorous manner, one should use formal mathematical logic.

Proof Using definition of set operations, we have

$$\begin{aligned} x \in (A \cup B) \cap C &\iff x \in A \cup B \text{ and } x \in C \\ &\iff (x \in A \text{ or } x \in B) \text{ and } x \in C \\ &\iff (x \in A \text{ and } x \in C) \text{ or } (x \in B \text{ and } x \in C) \\ &\iff (x \in A \cap C) \text{ or } (x \in B \cap C) \\ &\iff x \in (A \cap C) \cup (B \cap C) \end{aligned}$$

This means that every element of $(A \cup B) \cap C$ is also an element of $(A \cap C) \cup (B \cap C)$ and vice versa. Thus the two sets are equal. □

Remark For more than three subsets of U, observations obtained from Venn diagrams may not be correct. For four subsets, we need to draw 3-dimensional Venn diagrams.

Exercise 1.1

1. Let $A = \{x \in U : x \leq 10\}$, $B = \{x \in U : x \text{ is a prime number}\}$ and $C = \{x \in U : x \text{ is an even number}\}$, where $U = \{1, 2, 3, \ldots, 19\}$ is the universal set. Find the following sets.

(a)	$A \cap B$	(b)	$A \cap C$
(c)	$B \cap C$	(d)	$A \cup B$
(e)	$A \cup C$	(f)	$B \cup C$
(g)	$A \cup B \cup C$	(h)	$A \cap B \cap C$
(i)	$(A \cup B) \cap C$	(j)	$(A \cap B) \cup C$
(k)	$A \cap B'$	(l)	$A' \cap B'$

*2. Let A, B and C be subsets of a universal set U. For each of the following statements, determine whether it is true or not.

(a) $A - B = A' \cap B$

(b) $(A \cup B) \cap C = A \cup (B \cap C)$

(c) $(A' \cup B') \cap B = B - A$

A statement above is true means that it is true for all possible choices of A, B, C and U. To show that the statement is false, it is enough to give a counterexample. To show that it is true, you can draw a Venn diagram to convince yourself; but to be more rigorous, you should use formal mathematical logic.

1.2 Real Numbers

1.2.1 The Number Systems

(1) The numbers $0, 1, 2, 3, \ldots$ are called *natural numbers*. The set of all natural numbers is denoted by $\mathbb{N}$, that is,

$$\mathbb{N} = \{0, 1, 2, 3, \ldots\}.$$

Remark The three dots "..." means that the pattern is repeated indefinitely.

FAQ In some books, $\mathbb{N}$ is defined to be $\{1, 2, 3, \ldots\}$. Which one should we follow?

Answer Some authors do not include 0 in $\mathbb{N}$. This is just a convention; once we know the definition, it will not cause any problem. □

(2) The numbers $0, 1, -1, 2, -2, \ldots$ are called *integers*. The set of all integers is denoted by $\mathbb{Z}$, that is,

$$\mathbb{Z} = \{\cdots, -3, -2, -1, 0, 1, 2, 3, \cdots\}.$$

(3) Numbers in the form $\frac{p}{q}$ where $p, q \in \mathbb{Z}$ and $q \neq 0$ are called *rational numbers*. The set of all rational numbers is denoted by $\mathbb{Q}$, that is,

$$\mathbb{Q} = \left\{\frac{p}{q} : p, q \text{ are integers, and } q \neq 0\right\}.$$

Note $\mathbb{Z} \subseteq \mathbb{Q}$, that is, every integer is a rational number. For example, the integer 2 can be written as $\frac{2}{1}$ and is therefore a rational number.

All rational numbers can be represented by decimal numbers that *terminate*, such as $\frac{3}{4} = 0.75$, or by *non-terminating* but *repeating* decimals, such as $\frac{4}{11} = 0.363636\cdots$.

Numbers that can be represented by *non-terminating* and *non-repeating* decimals are called *irrational numbers*. For example, π and $\sqrt{2}$ are irrational numbers. The following shows the first 50 decimals of π:

$$\pi = 3.14159265358979323846264338327950288419716939937511\ldots$$

Remark The proof for the fact that π is irrational is difficult.

(4) Rational numbers together with irrational numbers are called *real numbers*. The set of all real numbers is denoted by $\mathbb{R}$.

In $\mathbb{R}$, we have the *algebraic operations* $+, \times$ (and $-, \div$ also) as well as *binary relations* $<, \leq, >, \geq$. Numbers greater than (respectively smaller than) 0 are called *positive* (respectively *negative*).

Real Number Line Real numbers can be represented by points on a line, called the *real number line*.

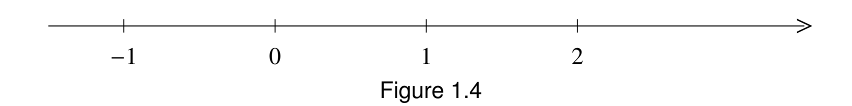


Figure 1.4

Notation The following nine types of subsets of $\mathbb{R}$ are called *intervals*:

$$[a,b] = \{x \in \mathbb{R} : a \leq x \leq b\} \tag{1.2.1}$$

$$(a,b) = \{x \in \mathbb{R} : a < x < b\} \tag{1.2.2}$$

$$[a,b) = \{x \in \mathbb{R} : a \leq x < b\} \tag{1.2.3}$$

$$(a,b] = \{x \in \mathbb{R} : a < x \leq b\} \tag{1.2.4}$$

$$[a,\infty) = \{x \in \mathbb{R} : a \leq x\} \tag{1.2.5}$$

$$(a,\infty) = \{x \in \mathbb{R} : a < x\} \tag{1.2.6}$$

$$(-\infty,b] = \{x \in \mathbb{R} : x \leq b\} \tag{1.2.7}$$

$$(-\infty,b) = \{x \in \mathbb{R} : x < b\} \tag{1.2.8}$$

$$(-\infty,\infty) = \mathbb{R} \tag{1.2.9}$$

where a and b are real numbers with $a < b$ and ∞ and $-\infty$ (read "*infinity*" and "*minus infinity*") are just symbols but not real numbers.

FAQ What are the meaning of ∞ and $-\infty$?

Answer Intuitively, you may imagine that there is a point, denoted by ∞, very far away on the right (and $-\infty$ on the left). So (a,∞) is the set whose elements are the points between a and ∞, that is, real numbers greater than a. □

Remark The notation (a,b), where $a < b$, has two different meanings. It denotes an ordered pair as well as an interval. To avoid ambiguity, some authors use $]a,b[$ to denote the open interval $\{x \in \mathbb{R} : a < x < b\}$. In this course, we will not use this notation. Readers can determine the meaning from the context.

Terminology

- Intervals in the form (a,b), $[a,b]$, $(a,b]$ and $[a,b)$ are called *bounded intervals* and those in the form $(-\infty,b)$, $(-\infty,b]$, (a,∞), $[a,\infty)$ and $(-\infty,\infty)$ are called *unbounded intervals*.
- Intervals in the form (a,b), $(-\infty,b)$, (a,∞) and $(-\infty,\infty)$ are called *open intervals*. For each of such intervals, the endpoint(s), if there is any, does not belong to the interval.
- Intervals in the form $[a,b]$, $(-\infty,b]$, $[a,\infty)$ and $(-\infty,\infty)$ are called *closed intervals*. For each of such intervals, the endpoint(s), if there is any, belongs to the interval.
- Intervals in the form $[a,b]$ are called *closed and bounded intervals*.

- A set $\{a\}$ with exactly one element of $\mathbb{R}$ is called a *degenerated interval* (its length is 0).
- Some authors also include $\emptyset$ as an interval (called the *empty interval*).

In this course, an interval means a nonempty, non-degenerated interval, that is, an infinite subset of $\mathbb{R}$ that can be written in the form (1.2.1), (1.2.2), (1.2.3), (1.2.4), (1.2.5), (1.2.6), (1.2.7), (1.2.8) or (1.2.9).

Example For each of the following pairs of intervals A and B,

(1) $A = [1, 5]$ and $B = (3, 10]$

(2) $A = [-2, 3]$ and $B = (7, 11]$

(3) $A = [-7, -2)$ and $B = [-2, \infty)$

- determine whether it is (i) an open interval, (ii) a closed interval , (iii) a bounded interval;
- find $A \cap B$ and determine whether it is an interval.
- find $A \cup B$ and determine whether it is an interval.

Solution

(1) Both A and B are not open intervals.
A is a closed interval but B is not a closed interval.
Both A and B are bounded intervals.
$A \cap B = (3, 5]$; it is an interval.
$A \cup B = [1, 10]$; it is an interval.

(2) Both A and B are not open intervals.
A is a closed interval but B is not a closed interval.
Both A and B are bounded intervals.
$A \cap B = \emptyset$; it is not an interval.
$A \cup B = [-2, 3] \cup (7, 11]$; it is not an interval.

(3) Both A and B are not open intervals.
B is a closed interval but A is not a closed interval.
A is a bounded interval but B is not a bounded interval.
$A \cap B = \emptyset$; it is not an interval.
$A \cup B = [-7, \infty)$; it is an interval.

□

1.2.2 Radicals

Definition

(1) Let a and b be real numbers and let q be a positive integer. If $a^q = b$, we say that a is a *qth root* of b.

Example

(a) -2 is the cube root of -8.

(b) 3 and -3 are the square roots of 9.

Note

(a) If q is odd, then every real number has a unique qth root.

(b) If q is even, then

  (i) every positive real number has two qth roots;

  (ii) negative real numbers do not have qth root;

  (iii) the qth root of 0 is 0.

(2) Let b be a real number and let q be a positive integer. The *principal qth root* of b, denoted by $\sqrt[q]{b}$, is defined as follows:

(a) if q is odd, $\sqrt[q]{b}$ is the unique qth root of b;

(b) if q is even,

  (i) $\sqrt[q]{b}$ is the positive qth root of b if $b > 0$;

  (ii) $\sqrt[q]{b}$ is undefined b if $b < 0$;

  (iii) $\sqrt[q]{b}$ is 0 if $b = 0$.

When $q = 2$, $\sqrt[2]{x}$ is simply written as $\sqrt{x}$.

FAQ Can we write $\sqrt{4} = \pm 2$?

Answer According to the definition, $\sqrt{4}$ is the principle square root of 4, which is the positive real number whose square is 4. That is, $\sqrt{4} = 2$. □

FAQ In solving $x^2 = 4$, we get $x = \pm 2$. Is this different from the above question?

Answer To find $\sqrt{4}$ is different from solving $x^2 = 4$.

(a) $\sqrt{4}$ is a uniquely defined real number.

(b) To solve $x^2 = 4$ is to find real numbers whose square is 4. There are two such numbers, namely 2 and -2.

Don't mix up the two questions. □

Example

(a) $\sqrt[4]{81} = 3$

(b) $\sqrt[3]{-8} = -2$

(c) $\sqrt{25} = \sqrt[2]{25} = 5$

(d) $\sqrt{0} = \sqrt[2]{0} = 0$

(e) $\sqrt[6]{-3}$ is undefined.

Terminology The symbol $\sqrt[q]{b}$ is called a *radical* (q is called the *index* and b the *radicand*).

FAQ Is $\sqrt{a^2} = a$ always true?

Answer It is true if (and only if) $a \geq 0$. If $a < 0$, we have $\sqrt{a^2} = -a$. □

(3) Let b be a positive real number. Let p and q be integers where $q > 0$. We define

$$b^{\frac{p}{q}} = \sqrt[q]{b^p},$$

which is the same as $\left(\sqrt[q]{b}\right)^p$.

Example $8^{\frac{2}{3}} = \sqrt[3]{8^2} = \sqrt[3]{64} = 4$

Remark Equivalently, we have $8^{\frac{2}{3}} = \left(\sqrt[3]{8}\right)^2 = 2^2 = 4$.

FAQ Are the rules for exponents on page 1 valid if m and n are rational numbers?

Answer The rules remain valid for rational exponents, provided that the base is positive (this is required in the definition of $b^{\frac{p}{q}}$). For example, we have $b^s b^t = b^{s+t}$, where $b > 0$ and $s, t \in \mathbb{Q}$.

Proof Write $s = \frac{m}{n}$ and $t = \frac{p}{q}$ where m, n, p, q are integers with $q, n > 0$. Note that

$$s = \frac{mq}{nq}, \qquad t = \frac{np}{nq} \qquad \text{and} \qquad s + t = \frac{mq + np}{nq}.$$

By definition (equivalent form), we have

$$b^s = \left(\sqrt[nq]{b}\right)^{mq} \qquad \text{and} \qquad b^t = \left(\sqrt[nq]{b}\right)^{np}.$$

Denote $\alpha = \sqrt[nq]{b}$. Then we have

$$\begin{aligned} b^s \cdot b^t &= \alpha^{mq} \cdot \alpha^{np} \\ &= \alpha^{mq+np} \\ &= \alpha^{nq(s+t)} \\ &= (\alpha^{nq})^{s+t} \\ &= b^{s+t} \end{aligned}$$

□

FAQ Can we define b raising to an irrational power? For example, can we define 2^π? How?

Answer This is deep question. The idea will be discussed Chapter 8. □

Exercise 1.2

1. Find the following sets.
   (a) $\{x \in \mathbb{R} : x^2 = 2\}$
   (b) $\{x \in \mathbb{R} : x \geq 0 \text{ and } x^2 = 2\}$
   (c) $\{x \in \mathbb{Q} : x^2 = 2\}$
2. Let $A = [1, 5]$, $B = [3, 9)$, $C = \{1, 5\}$ and $D = [5, \infty)$. Find
   (a) $A \cap B$ (b) $A \cup B$
   (c) $A - C$ (d) $B \cap C$
   (e) $C - B$ (f) $B - C$
   (g) $B - (B - C)$ (h) $A \cup D$
   (i) $C \cap D$

1.3 Solving Inequalities

An inequality in one unknown x can be written in one of the following forms:

(1) $F(x) > 0$

(2) $F(x) \geq 0$

(3) $F(x) < 0$

(4) $F(x) \leq 0$

where F is a function from a subset of $\mathbb{R}$ into $\mathbb{R}$.

Definition Consider an inequality in the form $F(x) > 0$ (the other cases can be treated similarly).

(1) A real number x_0 satisfying $F(x_0) > 0$ is called a *solution* to the inequality.

(2) The set of all solutions to the inequality is called the *solution set* to the inequality.

To solve an inequality means to find all the solutions to the inequality, or equivalently, to find the solution set.

In this section, we consider *polynomial inequalities*

$$a_n x^n + a_{n-1}x^{n-1} + \cdots + a_1 x + a_0 < 0 \quad (\text{or } > 0, \text{ or } \leq 0, \text{ or } \geq 0) \tag{1.3.1}$$

where $n \geq 1$ and $a_n \neq 0$.

When $n = 1$, (1.3.1) is a linear inequality. A revision for solving linear inequalities is given in Chapter 0. In the following examples, we consider several linear inequalities simultaneously.

Example Find the solution set to the following compound inequality:

$$1 \leq 3 - 2x \leq 9$$

Solution The inequality means

$$1 \leq 3 - 2x \qquad \text{and} \qquad 3 - 2x \leq 9.$$

Solving them separately, we get

$$\begin{array}{rcl} 2x & \leq & 2 \\ x & \leq & 1 \end{array} \qquad \text{and} \qquad \begin{array}{rcl} -6 & \leq & x \\ -3 & \leq & x. \end{array}$$

The solution set is $\{x \in \mathbb{R} : x \leq 1 \text{ and } -3 \leq x\} = \{x \in \mathbb{R} : -3 \leq x \leq 1\}$. □

Remark Using interval notation, the solution set can be written as $[-3, 1]$.

Example Find the solution set to the following:

$$2x + 1 < 3 \quad \text{and} \quad 3x + 10 < 4.$$

Give your answer using interval notation.

Solution Solving the inequalities separately, we get

$$\begin{array}{rcl} 2x & < & 2 \\ x & < & 1 \end{array} \qquad \text{and} \qquad \begin{array}{rcl} 3x & < & -6 \\ x & < & -2. \end{array}$$

Therefore, we have

$$\begin{aligned} \text{solution set} &= \{x \in \mathbb{R} : x < 1 \text{ and } x < -2\} \\ &= \{x \in \mathbb{R} : x < -2\} \\ &= (-\infty, -2). \end{aligned}$$

□

Example Find the solution set to the following:

$$2x + 1 > 9 \quad \text{and} \quad 3x + 4 < 10$$

Solution Solving the inequalities separately, we get

$$\begin{aligned} 2x &> 8 \\ x &> 4 \end{aligned} \quad \text{and} \quad \begin{aligned} 3x &< 6 \\ x &< 2. \end{aligned}$$

The solution set is $\{x \in \mathbb{R} : x > 4 \text{ and } x < 2\} = \emptyset$. □

1.3.1 Quadratic Inequalities

A quadratic inequality (in one unknown) is an inequality that can be written in the form

$$ax^2 + bx + c < 0 \quad (\text{or } > 0, \text{ or } \le 0, \text{ or } \ge 0) \tag{1.3.2}$$

where $a \neq 0$. This corresponds to $n = 2$ in (1.3.1).

We use an example to describe three methods for solving quadratic inequalities. The first two methods make use of the following properties of real numbers.

(1) $\alpha > 0$ and $\beta > 0 \implies \alpha \cdot \beta > 0$

(2) $\alpha < 0$ and $\beta < 0 \implies \alpha \cdot \beta > 0$

(3) $\alpha > 0$ and $\beta < 0 \implies \alpha \cdot \beta < 0$

From these we get

(4) $\alpha \cdot \beta > 0 \iff (\alpha > 0 \text{ and } \beta > 0) \text{ or } (\alpha < 0 \text{ and } \beta < 0)$

(5) $\alpha \cdot \beta < 0 \iff (\alpha > 0 \text{ and } \beta < 0) \text{ or } (\alpha < 0 \text{ and } \beta > 0)$

Example Find the solution set to the inequality $x^2 + 2x - 15 > 0$.

Solution

(*Method 1*) First we factorize the quadratic polynomial:

$$\begin{aligned} x^2 + 2x - 15 &> 0 \\ (x + 5)(x - 3) &> 0, \end{aligned}$$

and then apply Property (4):

$$\begin{aligned} (x + 5 > 0 \text{ and } x - 3 > 0) &\quad \text{or} \quad (x + 5 < 0 \text{ and } x - 3 < 0) \\ (x > -5 \text{ and } x > 3) &\quad \text{or} \quad (x < -5 \text{ and } x < 3) \\ x > 3 &\quad \text{or} \quad x < -5 \end{aligned}$$

The solution set is $\{x \in \mathbb{R} : x < -5 \text{ or } x > 3\} = (-\infty, -5) \cup (3, \infty)$.

(*Method 2*) By factorization, we have

$$L.S. = (x+5)(x-3).$$

The left-side is zero when $x = -5$ or 3. These two points divide the real number line into three intervals:

$$(-\infty, -5), \qquad (-5, 3), \qquad (3, \infty).$$

In the following table, the first two rows give the signs of $(x + 5)$ and $(x - 3)$ on each of these intervals. Hence, using Properties (1), (2) and (3), we obtain the signs of $(x + 5)(x - 3)$ in the third row.

	$x < -5$	$x = -5$	$-5 < x < 3$	$x = 3$	$x > 3$
$x + 5$	$-$	0	$+$	$+$	$+$
$x - 3$	$-$	$-$	$-$	0	$+$
$(x + 5)(x - 3)$	$+$	0	$-$	0	$+$

The solution set is $(-\infty, -5) \cup (3, \infty)$.

Remark To determine the sign of $(x + 5)$, first we note that it is 0 when $x = -5$. Since $(x + 5)$ increases as x increases, it is positive when $x > -5$ and negative when $x < -5$.

(*Method 3*) The graph of $y = x^2 + 2x - 15$ is a parabola opening upward and it cuts the x-axis at $x_1 = -5$ and $x_2 = 3$. To solve the inequality $x^2 + 2x - 15 > 0$ means to find all x such that the corresponding points on the parabola has y-coordinates greater than 0. From the graph, we see that the parabola is above the x-axis if and only if $x < -5$ or $x > 3$. Therefore, the solution set is $(-\infty, -5) \cup (3, \infty)$.

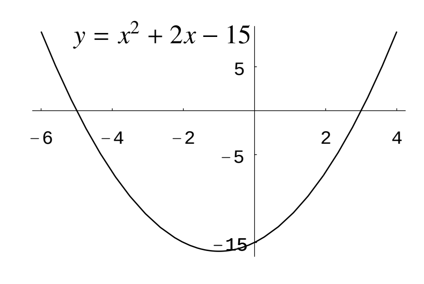


Figure 1.5

□

1.3.2 Polynomial Inequalities with degrees ≥ 3

In this section, we consider polynomial inequalities (1.3.1) of degree $n \geq 3$. To solve such polynomial inequalities, for example $p(x) > 0$, we can use methods similar to that for quadratic inequalities. The first step is to factorize $p(x)$.

Example Factorize the polynomial $p(x) = x^3 + 3x^2 - 4x - 12$.

Solution First we try to find a factor of the form $(x - c)$ where c is an integer. For this, we try

$$c = \pm 1, \pm 2, \pm 3, \pm 4, \pm 6, \pm 12.$$

Direct substitution gives $p(2) = 0$ and so $(x - 2)$ is a factor of $p(x)$. Using long division, we obtain

$$x^3 + 3x^2 - 4x - 12 = (x - 2)(x^2 + 5x + 6)$$

and then using inspection we get

$$x^3 + 3x^2 - 4x - 12 = (x - 2)(x + 2)(x + 3).$$

□

In the above procedure, we make use of the following

Theorem 1.3.1 *Let*

$$p(x) = c_n x^n + c_{n-1}x^{n-1} + \cdots + c_1 x + c_0$$

be a polynomial of degree n where $c_0, c_1, \ldots, c_n \in \mathbb{Z}$. Suppose $(ax - b)$ is a factor of $p(x)$ where $a, b \in \mathbb{Z}$. Then a divides c_n and b divides c_0.

FAQ Can we use Factor Theorem to find all the linear factors?

Answer If some linear factors are repeated more than once, we can't determine which one is repeated (and also how many times?). For example, let $p(x) = x^3 - 3x + 2$. Using Factor Theorem, we get linear factors $(x - 1)$ and $(x + 2)$. It is incorrect to write $p(x) = (x - 1)(x + 2)$. Indeed, we have

$$p(x) = (x - 1)^2(x + 2).$$

Remark We say that $(x - 1)$ is a factor of $p(x)$ repeated twice. □

Example Find the solution set to the inequality $x^3 + 3x^2 - 4x - 12 \le 0$.

Solution Factorizing the polynomial $p(x)$ on the left side we obtain

$$p(x) = x^3 + 3x^2 - 4x - 12 = (x - 2)(x + 2)(x + 3).$$

The sign of $p(x)$ can be determined from the following table:

	$x < -3$	$x = -3$	$-3 < x < -2$	$x = -2$	$-2 < x < 2$	$x = 2$	$2 < x$
$x - 2$	$-$	$-$	$-$	$-$	$-$	0	+
$x + 2$	$-$	$-$	$-$	0	+	+	+
$x + 3$	$-$	0	+	+	+	+	+
$p(x)$	$-$	0	+	0	$-$	0	+

The solution set is $\{x \in \mathbb{R} : x \le -3 \text{ or } -2 \le x \le 2\} = (-\infty, -3] \cup [-2, 2]$. □

FAQ Can we use *Method 1* described in Section 1.3.1?

Answer You can use that method. However the "and/or" logic is more complicated. If the degree of the polynomial is 3, there are 4 cases; if the degree is 4, there are 8 cases. The number of cases doubles if the degree increases by 1.

For the table method, if the degree increases by 1, the number of factors and the number of intervals increase by (at most) 1 only. □

FAQ Can we use graphical method?

Answer If you know the graph of $y = x^3 + 3x^2 - 4x - 12$, you can write down the solution immediately. For that, you need to know the x-intercepts (obtained by factoring the polynomial) and also the shape of the graph (on which interval is the graph going up or down?). This will be discussed in Chapter 5.

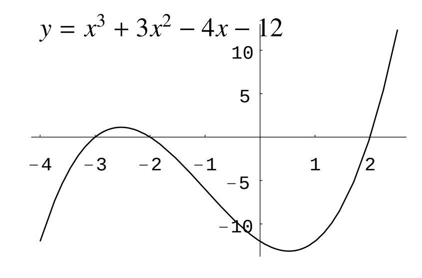


Figure 1.6 □

Exercise 1.3

1. Solve the following inequalities:

   (a) $2x - 3 \geq 4 + 7x$ (b) $8(x + 1) - 2 < 5(x - 6) + 7$

   (c) $\dfrac{x^2 - 3x + 7}{x^2 + 1} < 1$ (d) $(2x + 7)(5 - 11x) \leq 0$

   (e) $x^2 - 2x - 3 < 0$ (f) $2x^2 - 3x > 4$

   (g) $2x^2 - 3x < -4$ (h) $\dfrac{2x + 3}{x - 4} \geq 0$

   (i) $\dfrac{2x + 3}{x - 4} < 1$

   Note: $\frac{a}{b} < 0$ is equivalent to $a \cdot b < 0$.

2. Factorize the following polynomials:

   (a) $2x^3 + 7x^2 - 15x$ (b) $2x^3 + 3x^2 - 2x - 3$

   (c) $x^3 - x^2 - x - 2$ (d) $x^4 - 3x^3 - 13x^2 + 15x$

   (e) $x^4 - 3x^3 + x^2 + 3x - 2$ (f) $x^4 - x^3 + x^2 - 3x + 2$

3. Solve the following inequalities:

   (a) $(x - 4)(9 - 5x)(2x + 3) < 0$ (b) $(x - 3)(2x + 1)^2 \leq 0$

   (c) $x^3 - 2x^2 - 5x + 6 < 0$ (d) $-2x^3 + x^2 + 15x - 18 \leq 0$

   (e) $x^3 - x^2 - 5x - 3 > 0$ (f) $x^3 + 3x^2 + 5x + 3 \leq 0$

   (g) $x^4 + 2x^3 - 13x^2 - 14x + 24 > 0$ (h) $6x^4 + x^3 - 15x^2 \leq 0$

Chapter 2

Functions and Graphs

2.1 Functions

Informal definition Let A and B be sets. A *function from A into B*, denoted by $f : A \longrightarrow B$, is a "*rule*" that assigns to each element of A exactly one element of B.

Remark If the sets A and B are understood (or are not important for the problem under consideration), instead of saying "*a function from A into B*", we simply say "*a function*" and instead of writing $f : A \longrightarrow B$, we simply write f.

Terminology and Notation The sets A and B are called the *domain* and *codomain* of f respectively. The domain of f is denoted by $\text{dom}(f)$.

FAQ In the above informal definition, what is the meaning of a *rule*?

Answer It is difficult to tell what a rule is. The above informal definition describes the idea of a *function*. There is a rigorous definition. However, it involves more definitions and notations. Interested readers may consult books on set theory or foundation of mathematics. □

Notation & Terminology Let f be a function. For each x belonging to the domain of f, the corresponding element (in the codomain of f) assigned by f is denoted by $f(x)$ and is called the *image* of x under f.

Remark Some people write $f(x)$ to denote a function. This notation may be misleading because it also means an image. However, sometimes for convenience, such notations are used. For example, we write x^2 to denote the square function, that is, the function f (from $\mathbb{R}$ into $\mathbb{R}$) given by $f(x) = x^2$.

In this course, most of the functions we consider are functions whose domains and codomains are subsets of $\mathbb{R}$. A variable that represents the "*input numbers*" for a function is called an *independent* variable. A variable that represents the "*output numbers*" is called a *dependent* variable because its value depends on the value of the independent variable.

Example Consider the function $f : \mathbb{R} \longrightarrow \mathbb{R}$ given by

$$f(x) = x^2 + 2.$$

We may also write $y = x^2 + 2$ to represent this function. For each input x, the function gives exactly one output $x^2 + 2$, which is y. If $x = 3$, then $y = 11$; if $x = 6$, then $y = 38$ etc. The independent variable is x and the dependent variable is y.

Example Let $g(x) = x^2 - 3x + 7$. Find the following:

(1) $g(10)$

(2) $g(a + 1)$

(3) $g(r^2)$

(4) $g(x + h)$

(5) $\dfrac{g(x+h) - g(x)}{h}$

Solution

(1)
$$\begin{aligned} g(10) &= 10^2 - 3(10) + 7 \\ &= 77 \end{aligned}$$

(2)
$$\begin{aligned} g(a+1) &= (a+1)^2 - 3(a+1) + 7 \\ &= (a^2 + 2a + 1) - 3a - 3 + 7 \\ &= a^2 - a + 5 \end{aligned}$$

(3)
$$\begin{aligned} g(r^2) &= (r^2)^2 - 3(r^2) + 7 \\ &= r^4 - 3r^2 + 7 \end{aligned}$$

(4)
$$\begin{aligned} g(x+h) &= (x+h)^2 - 3(x+h) + 7 \\ &= x^2 + 2xh + h^2 - 3x - 3h + 7 \end{aligned}$$

(5)
$$\begin{aligned} \frac{g(x+h) - g(x)}{h} &= \frac{[(x+h)^2 - 3(x+h) + 7] - (x^2 - 3x + 7)}{h} \\ &= \frac{(x^2 + 2xh + h^2 - 3x - 3h + 7) - (x^2 - 3x + 7)}{h} \\ &= \frac{2xh + h^2 - 3h}{h} \\ &= 2x + h - 3 \end{aligned}$$

□

Exercise 2.1

1. Let $f(x) = \frac{x-5}{x^2+4}$. Find the following:

   (a) $f(2)$ (b) $f(3.5)$

   (c) $f(a + 1)$ (d) $f(\sqrt{a})$

   (e) $f(a^2)$ (f) $f(a) + f(1)$

2. Let $f(x) = \frac{x}{x+1}$ and $g(x) = \sqrt{x-1}$. Find the following:

   (a) $f(1) + g(1)$ (b) $f(2)g(2)$

   (c) $\frac{f(3)}{g(3)}$ (d) $f(a - 1) + g(a + 1)$

   (e) $f(a^2 + 1)g(a^2 + 1)$

3. Let $f(x) = x^2 - 3x + 4$. Find and simplify the following:

   (a) $f(a+b)$ (b) $\frac{f(1+h) - f(1)}{h}$

   (c) $\frac{f(a+h) - f(a)}{h}$

2.2 Domains and Ranges of Functions

To describe a function f, sometimes we just write down the rule defining f, omitting its domain and codomain.

- In this course, the codomain is always taken to be $\mathbb{R}$ unless otherwise stated.
- For the domain, it can be determined from the rule defining the function. For example, $f(x) = \sqrt{x}$ is defined for all real numbers $x \geq 0$ but undefined for $x < 0$. Therefore, we may take $[0, \infty)$ as the domain of f. The domain obtained in this way is called the *natural domain* of the function. For a function that is described by formula, we always take its domain to be the natural domain unless otherwise stated.

Summary Suppose f is a function described by a formula. Then the domain of f is the set of all real numbers x such that $f(x)$ is defined.

Remark For functions appeared in many applied problems, we do not take their natural domains. For example, the area A of a circle (to be more accurate, a circular region) with radius r is given by $A(r) = \pi r^2$. Although πr^2 is defined for all real numbers r, for the area function A, its domain is taken to be $\{r \in \mathbb{R} : r > 0\} = (0, \infty)$.

FAQ For the above area function, can we take the domain to be $[0, \infty)$?

Answer When the radius is 0, we get a point only. A point may be considered to be a circle, called a *degenerated circle*. Under this convention, 0 is included in the domain. In many problems, it doesn't matter whether we take $(0, \infty)$ or $[0, \infty)$ as the domain. □

Example For each of the following functions, find its (natural) domain.

(1) $f(x) = x^2 + 3$;

(2) $g(x) = \dfrac{1}{x-2}$;

(3) $h(x) = \sqrt{1 + 5x}$

Solution

(1) Since $f(x) = x^2 + 3$ is defined for all real numbers x, the domain of f is $\mathbb{R}$.

(2) Note that $g(x)$ is defined for all real numbers x except 2.
The domain of g is $\{x \in \mathbb{R} : x \neq 2\} = \mathbb{R} \setminus \{2\}$.
Remark The domain can also be written as $\{x \in \mathbb{R} : x < 2 \text{ or } x > 2\} = (-\infty, 2) \cup (2, \infty)$.

(3) Note that $\sqrt{1 + 5x}$ is defined if and only if $1 + 5x \geq 0$.
The domain of h is
$$\{x \in \mathbb{R} : 1 + 5x \geq 0\} = \{x \in \mathbb{R} : x \geq -\tfrac{1}{5}\} = [-\tfrac{1}{5}, \infty).$$
□

Definition Let $f : A \longrightarrow B$ be a function and let $S \subseteq A$. The *image* of S under f, denoted by $f[S]$, is the subset of B given by

$$f[S] = \{y \in B : y = f(x) \text{ for some } x \in S\}.$$

Note $f[S]$ is the subset of B consisting of all the images under f of elements in S.

Example

(1) Let $f : \mathbb{R} \longrightarrow \mathbb{R}$ be the function given by $f(x) = x^2$. For $S = \{1, 2, 3\}$, we have $f[S] = \{1, 4, 9\}$.

(2) Let $f : \mathbb{R} \longrightarrow \mathbb{R}$ be the function given by $f(x) = 2x + 1$. For $S = [0, 1]$, we have $f[S] = [1, 3]$.

Definition Let $f : A \longrightarrow B$ be a function. The *range* of f, denoted by $\operatorname{ran}(f)$, is the image of A under f, that is, $\operatorname{ran}(f) = f[A]$.

Remark By definition, $\operatorname{ran}(f) = \{y \in B : y = f(x) \text{ for some } x \in A\}$. The condition

(∗) $y = f(x)$ for some $x \in A$

means that y is an output (image) corresponding to some input (element of A). When A and B are subsets of $\mathbb{R}$, (∗) means that the equation

$$y = f(x)$$

has at least one solution belonging to A.

Example Let $f : \mathbb{R} \longrightarrow \mathbb{R}$ be the function given by $f(x) = x^2 + 2$. Then

(1) 3 belongs to the range of f because $f(1) = 3$, that is, 3 is the image of 1 under f. In terms of solving equation, 3 belongs to the range means that the equation $3 = x^2 + 1$ has solution in $\mathbb{R}$ (the domain of f). Indeed, the equation has two solutions in $\mathbb{R}$, namely 1 and -1;

(2) 2 belongs to the range because the equation $2 = x^2 + 2$ has solution in $\mathbb{R}$, namely, 0;

(3) 1 does not belong to the range because the equation $1 = x^2 + 2$ has no solution in $\mathbb{R}$.

Steps to find range of function To find the range of a function f described by formula, where the domain is taken to be the natural domain:

(1) Put $y = f(x)$.

(2) Solve x in terms of y.

(3) The range of f is the set of all real numbers y such that x can be solved.

Example For each of the following functions, find its range.

(1) $f(x) = x^2 + 2$

(2) $g(x) = \dfrac{1}{x - 2}$

(3) $h(x) = \sqrt{1 + 5x}$

Solution

(1) Put $y = f(x) = x^2 + 2$.

Solve for x.
$$\begin{aligned} x^2 &= y - 2 \\ x &= \pm\sqrt{y-2}. \end{aligned}$$

Note that x can be solved if and only if $y - 2 \geq 0$.

The range of f is $\{y \in \mathbb{R} : y - 2 \geq 0\} = \{y \in \mathbb{R} : y \geq 2\} = [2, \infty)$.

Alternatively, to see that the range is $[2, \infty)$, we may use the graph of $y = x^2 + 2$ which is a parabola. The lowest point (vertex) is $(0, 2)$. For any $y \geq 2$, we can always find $x \in \mathbb{R}$ such that $f(x) = y$.

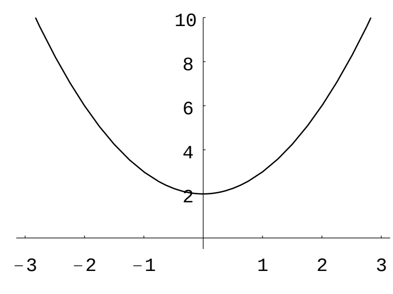


Figure 2.1

(2) Put $y = g(x) = \dfrac{1}{x-2}$.

Solve for x.
$$\begin{aligned} y &= \frac{1}{x-2} \\ x - 2 &= \frac{1}{y} \\ x &= \frac{1}{y} + 2. \end{aligned}$$

Note that x can be solved if and only if $y \neq 0$.

The range of g is $\{y \in \mathbb{R} : y \neq 0\} = \mathbb{R} \setminus \{0\}$.

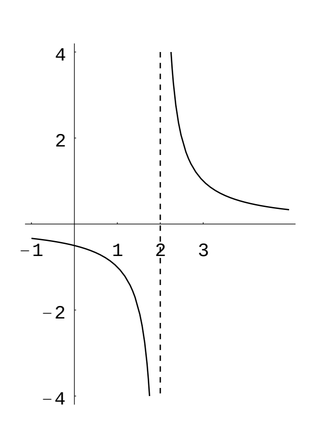


Figure 2.2

(3) Put $y = h(x) = \sqrt{1 + 5x}$. Note that y cannot be negative.

Solve for x.
$$\begin{aligned} y &= \sqrt{1+5x}, & y \geq 0 \\ y^2 &= 1 + 5x, & y \geq 0 \\ x &= \frac{y^2 - 1}{5}, & y \geq 0. \end{aligned}$$

Note that x can always be solved for every $y \geq 0$.

The range of h is $\{y \in \mathbb{R} : y \geq 0\} = [0, \infty)$.

Remark $y = \sqrt{1 + 5x} \Longrightarrow y^2 = 1 + 5x$ but the converse is true only if $y \geq 0$.

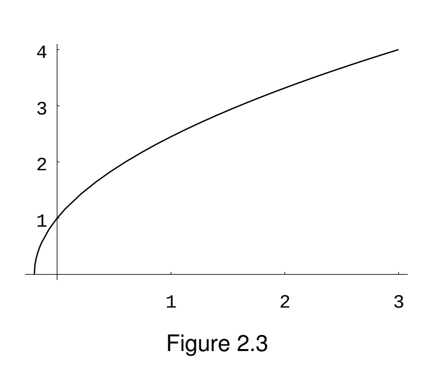


Figure 2.3

□

Example Let $f(x) = \sqrt{x + 7} - \sqrt{x^2 + 2x - 15}$. Find the domain of f.

Solution Note that $f(x)$ is defined if and only if $x + 7 \geq 0$ and $x^2 + 2x - 15 \geq 0$.
Solve the two inequalities separately:

- $$\begin{aligned} x + 7 &\geq 0 \\ x &\geq -7; \end{aligned}$$

- $$\begin{aligned} x^2 + 2x - 15 &\geq 0 \\ (x+5)(x-3) &\geq 0 \end{aligned}$$

	$x < -5$	$x = -5$	$-5 < x < 3$	$x = 3$	$x > 3$
$x - 3$	$-$	$-$	$-$	0	$+$
$x + 5$	$-$	0	$+$	$+$	$+$
$(x-3)(x+5)$	$+$	0	$-$	0	$+$

  thus, $x \leq -5$ or $x \geq 3$.

Therefore, we have
$$\begin{aligned} \text{dom}(f) &= \{x \in \mathbb{R} : x \geq -7 \text{ and } (x \leq -5 \text{ or } x \geq 3)\} \\ &= \{x \in \mathbb{R} : (x \geq -7 \text{ and } x \leq -5) \text{ or } (x \geq -7 \text{ and } x \geq 3)\} \\ &= \{x \in \mathbb{R} : -7 \leq x \leq -5 \text{ or } x \geq 3\} \\ &= [-7, -5] \cup [3, \infty). \end{aligned}$$

□

Example Let $f(x) = \dfrac{2x+1}{x^2+1}$. Find the range of f.

Solution

Put $y = f(x) = \frac{2x+1}{x^2+1}$.

Solve for x.
$$\begin{aligned} y &= \frac{2x+1}{x^2+1} \\ yx^2 + y &= 2x + 1 \\ yx^2 - 2x + (y-1) &= 0 \\ x &= \frac{2 \pm \sqrt{4 - 4y(y-1)}}{2y} \text{ if } y \neq 0, \qquad x = \frac{-1}{2} \text{ if } y = 0, \\ &= \frac{1 \pm \sqrt{1 - y^2 + y}}{y} \text{ if } y \neq 0. \end{aligned}$$

Combining the two cases, we see that x can be solved if and only if $1 - y^2 + y \geq 0$, that is, $y^2 - y - 1 \leq 0$.

The range of f is $\{y \in \mathbb{R} : y^2 - y - 1 \leq 0\}$.

To solve the inequality $y^2 - y - 1 \leq 0$, first we find the zero of the left-side by quadratic formula to get $\frac{1 \pm \sqrt{5}}{2}$.

By the factor theorem and comparing coefficient of y^2, we see that $y^2 - y - 1 = \left(y - \frac{1-\sqrt{5}}{2}\right)\left(y - \frac{1+\sqrt{5}}{2}\right)$

	$y < \frac{1-\sqrt{5}}{2}$	$y = \frac{1-\sqrt{5}}{2}$	$\frac{1-\sqrt{5}}{2} < y < \frac{1+\sqrt{5}}{2}$	$y = \frac{1+\sqrt{5}}{2}$	$y > \frac{1+\sqrt{5}}{2}$
$y - \frac{1-\sqrt{5}}{2}$	$-$	0	$+$	$+$	$+$
$y - \frac{1+\sqrt{5}}{2}$	$-$	$-$	$-$	0	$+$
$y^2 - y - 1$	$+$	0	$-$	0	$+$

From the table, we see that
$$\begin{aligned} \text{ran}(f) &= \{y \in \mathbb{R} : \frac{1-\sqrt{5}}{2} \leq y \leq \frac{1+\sqrt{5}}{2}\} \\ &= [\frac{1-\sqrt{5}}{2}, \frac{1+\sqrt{5}}{2}] \end{aligned}$$

□

Remark To solve the inequality $y^2 - y - 1 \le 0$, we may also use graphical method:

The figure shown is the graph of $z = y^2 - y - 1$, where the horizontal and vertical axes are the y-axis and the z-axis respectively.

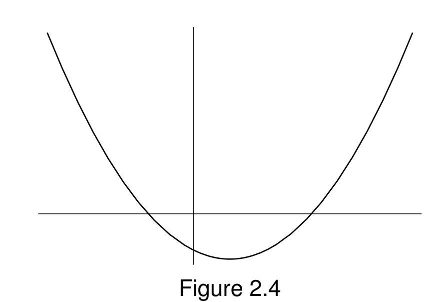


Figure 2.4

Exercise 2.2

1. For each of the following functions f, find its domain.

   (a) $f(x) = x^2 - 5$ (b) $f(x) = \frac{2}{5x+6}$

   (c) $f(x) = \frac{1}{x^2 - 5}$ (d) $f(x) = \frac{1}{x^2 - 2x - 3}$

   (e) $f(x) = \frac{1}{\sqrt{2x-3}}$ (f) $f(x) = \frac{1}{1-2x} - \sqrt{x+3}$

   (g) $f(x) = \frac{3}{1-x^2} + \sqrt{2x+5}$ (h) $f(x) = \frac{1}{\sqrt{x^2 + 3x - 10}}$

2. For each of the following functions f, find its range.

   (a) $f(x) = x^2 - 5$ (b) $f(x) = x^2 - 2x - 3$

   (c) $f(x) = \frac{2}{5x+6}$ (d) $f(x) = 3 - \frac{1}{2x-1}$

   (e) $f(x) = \frac{1}{\sqrt{2x-3}}$ (f) $f(x) = \frac{1}{x^2 - 5}$

   (g) $f(x) = \frac{1}{x^2 - 2x - 3}$

3. Consider a rectangle with perimeter 28 (units). Let the width of the rectangle be w (units) and let the area of the region enclosed by the rectangle be A (square units). Express A as a function of w. State the domain of A and find the range of A.

2.3 Graphs of Equations

Recall that an ordered pair of real numbers is denoted by (x_0, y_0) where x_0 and y_0 are real numbers. The set of all ordered pairs is denoted by $\mathbb{R}^2$ (read "*R two*"). The superscript 2 indicates that elements in $\mathbb{R}^2$ are represented by two real numbers. Since an ordered pair of real numbers represents a point in the coordinate plane, $\mathbb{R}^2$ can be identified with the plane.

Let $f : A \longrightarrow \mathbb{R}$ be a function where $A \subseteq \mathbb{R}^2$. Each element in the domain of f is an ordered pair (x, y) of real numbers. Its image under f is denoted by $f((x, y))$, or simply $f(x, y)$. Functions whose domains are subsets of $\mathbb{R}^2$ are called *functions of two variables*.

Example Let $f : \mathbb{R}^2 \longrightarrow \mathbb{R}$ be the function given by $f(x, y) = x + y^2$. Then we have

(1) $f(1, 2) = 1 + 2^2 = 5$

(2) $f(2, 1) = 2 + 1^2 = 3$

Consider an equation in the form

$$F(x, y) = 0 \tag{2.3.1}$$

where F is a function of two variables. The set of all ordered pairs (x, y) satisfying (2.3.1) is called the *graph* of (2.3.1). That is, the graph is the following subset of $\mathbb{R}^2$:

$$\{(x, y) \in \mathbb{R}^2 : F(x, y) = 0\}.$$

Since ordered pairs can be considered as points in the coordinate plane, the graph can be considered as a subset of the plane.

Example Consider the following equation

$$2x + 3y - 4 = 0.$$

(1) Since $2(2) + 3(0) - 4 = 0$, the point (ordered pair) $P(2, 0)$ belongs to the graph of the equation.

(2) Since $2(-1) + 3(2) - 4 = 0$, point $Q(-1, 2)$ belongs to the graph of the equation.

(3) Since $2(1) + 3(2) - 4 = 4 \neq 0$, the point $R(1, 2)$ does not belong to the graph.

Remark The graph of the equation is the line passing through P and Q.

Definition An *x-intercept* (respectively a *y-intercept*) of the graph of an equation $F(x, y) = 0$ is a point where the graph intersects the x-axis (respectively the y-axis).

Example The graph of the equation

$$2x + 3y - 4 = 0 \tag{2.3.2}$$

is a line. Its x-intercept is $(2, 0)$ and its y-intercept is $(0, \frac{4}{3})$. These are obtained by putting $y = 0$ and $x = 0$ respectively into (2.3.2).

Example The graph of the equation

$$x^2 + y^2 = 1$$

is a circle centered at the origin $(0, 0)$ with radius 1.

The graph has two x-intercepts, namely $(1, 0)$ and $(-1, 0)$ and two y-intercepts, namely, $(0, 1)$ and $(0, -1)$.

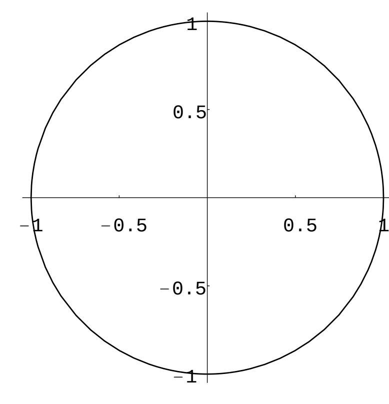


Figure 2.5

Example Find the x-intercept(s) and y-intercept(s) of the graph of

$$y = x^2 - 5x + 6 \tag{2.3.3}$$

Solution To find the x-intercepts, we put $y = 0$ in (2.3.3). Solving

$$\begin{aligned} 0 &= x^2 - 5x + 6 \\ 0 &= (x - 2)(x - 3) \end{aligned}$$

we get $x = 2$ or $x = 3$. Thus the x-intercepts are $(2, 0)$ and $(3, 0)$.

To find the y-intercept, we put $x = 0$ in (2.3.3) and get $y = 6$. Thus the y-intercept is $(0, 6)$. □

Symmetry Consider the graph of the equation

$$y = x^2.$$

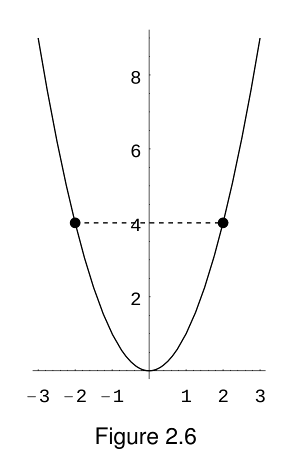


Figure 2.6

The graph is a parabola. If (a, b) is a point belonging to the parabola, that is $b = a^2$, then $(-a, b)$ also belongs to the parabola since $b = (-a)^2$. Note that

- the line segment joining (a, b) and $(-a, b)$ is perpendicular to the y-axis;
- the two points (a, b) and $(-a, b)$ are equidistant from the y-axis (distances to the y-axis are the same).

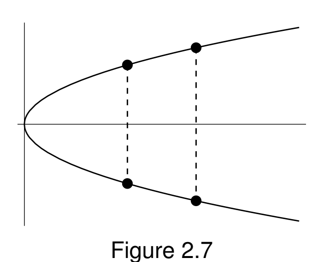
We say that the parabola is *symmetric* about the y-axis.

In general, a subset $\mathcal{A}$ of the plane is said to be *symmetric* about a line ℓ if the following condition is satisfied: For any point P belonging to $\mathcal{A}$ (but not belonging to ℓ), there is a point Q belonging to $\mathcal{A}$ such that

(1) the line segment PQ is perpendicular to ℓ;

(2) P and Q are equidistant from ℓ.

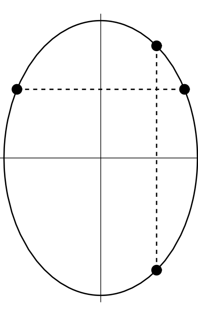
Example The parabola given by $x = y^2$ is symmetric about the x-axis.


Figure 2.7

Example The graph of $2x^2 + y^2 = 6$ is an ellipse. It is symmetric about the x-axis and also symmetric about the y-axis.
If (a, b) is a point belonging to the ellipse, then the point $(-a, -b)$ also belongs to the ellipse. Note that

- the line segment joining (a, b) and $(-a, -b)$ passes through the origin;
- the points (a, b) and $(-a, -b)$ are equidistant from the origin.

We say that the ellipse is symmetric about the origin.


Figure 2.8

In general, a subset $\mathcal{A}$ of the plane is said to be *symmetric* about a point C if the following condition is satisfied: For any point P belonging to $\mathcal{A}$ (but different from C), there is a point Q belonging to $\mathcal{A}$ such that

(1) the line segment PQ passes through C;

(2) P and Q are equidistant from C.

Example The graph of $y = x^3$ is symmetric about the origin.

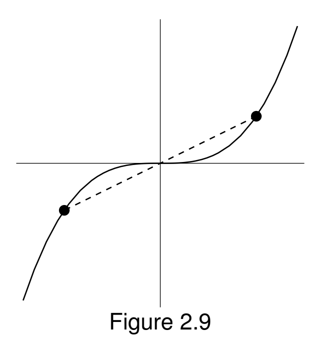


Figure 2.9

We close this section with the following example of finding intersection of two curves (in fact, one is a line). This is the same as solving a system of two equations in two unknowns.

Example Let $\mathcal{E}$ and $\mathcal{L}$ be the ellipse and the line given by

$$2x^2 + y^2 = 6 \quad \text{and} \quad x + 2y - 3 = 0$$

respectively. Find $\mathcal{E} \cap \mathcal{L}$.

Solution We need to solve the following system:

$$2x^2 + y^2 = 6 \qquad (2.3.4)$$

$$x + 2y - 3 = 0 \qquad (2.3.5)$$

From (2.3.5), we get $x = 3 - 2y$. Putting into (2.3.4) and solving

$$\begin{aligned} 2(3 - 2y)^2 + y^2 &= 6 \\ 2(9 - 12y + 4y^2) + y^2 &= 6 \\ 9y^2 - 24y + 12 &= 0 \\ 3(y - 2)(3y - 2) &= 0 \end{aligned}$$

we get $y = 2$ or $y = \frac{2}{3}$. Substitute back into (2.3.5), we get $(x, y) = (-1, 2)$ or $(\frac{5}{3}, \frac{2}{3})$.

Therefore we have $\mathcal{E} \cap \mathcal{L} = \{(-1, 2), (\frac{5}{3}, \frac{2}{3})\}$. □

Exercise 2.3

1. Consider the graph of $2x^2 + 3y^2 = 4$ (which is an ellipse). Find its x- and y-intercepts.
2. Suppose the graph of $y = ax^2 + bx + c$ has x-intercepts $(2, 0)$ and $(-3, 0)$ and y-intercept $(0, -6)$. Find a, b and c.
3. Consider the graph of $y = x^2 + 4x + 5$.
   (a) Find its x- and y-intercepts.
   (b) Show that the graph lies entirely above the x-axis.
4. Let $C = \{(x, y) \in \mathbb{R}^2 : x^2 + y^2 = 5\}$, $E = \{(x, y) \in \mathbb{R}^2 : x^2 + 2y^2 = 6\}$ and $L = \{(x, y) \in \mathbb{R}^2 : 2x + y - 3 = 0\}$. Find the following:
   (a) $L \cap C$ (b) $L \cap E$ (c) $C \cap E$
5. Let $C = \{(x, y) \in \mathbb{R}^2 : x^2 + y^2 = 1\}$ and $L = \{(x, y) \in \mathbb{R}^2 : ax + y = 2\}$ where a is a constant. Find the values of a such that $C \cap L$ is a singleton (that is, a set with only one element).

2.4 Graphs of Functions

Let $f : A \longrightarrow \mathbb{R}$ be a function where $A \subseteq \mathbb{R}$. The *graph* of f is the following subset of $\mathbb{R}^2$:

$$\{(x, y) \in \mathbb{R}^2 : x \in A \text{ and } y = f(x)\}.$$

Example

(1) **Constant Functions** A *constant function* is a function f that is given by

$$f(x) = c,$$

where c is a constant (a real number).

The domain of every constant function is $\mathbb{R}$.

The range is a singleton: $\{c\}$.

The graph is a horizontal line whose y-intercept is $(0, c)$.

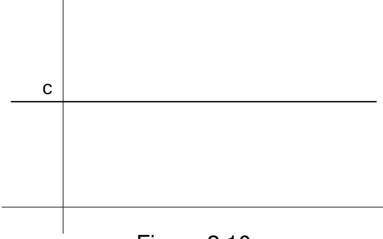


Figure 2.10

Remark Let $f(x) = x^0$. Note that for all $x \neq 0$, we have $f(x) = 1$ and that $f(0)$ is undefined. So there is a small difference between f and the constant function 1 whose domain is $\mathbb{R}$. However, for convenience, we treat the function x^0 as the constant function 1.

In the above discussion, we use the symbol 1 to represent the function with domain and codomain equal to $\mathbb{R}$ and assigning every $x \in \mathbb{R}$ to the number 1. Thus the symbol 1 has two different meanings. It may be a function or a number. This abuse of notation is sometimes used in mathematics. Readers can determine the meaning from the context.

(2) **Linear Functions** A *linear function* is a function f given by

$$f(x) = ax + b,$$

where a and b are constants and $a \neq 0$.

The domain of every linear function is $\mathbb{R}$.

The range is also $\mathbb{R}$ (note that a is assumed to be non-zero).

The graph is a line with slope a and y-intercept $(0, b)$.

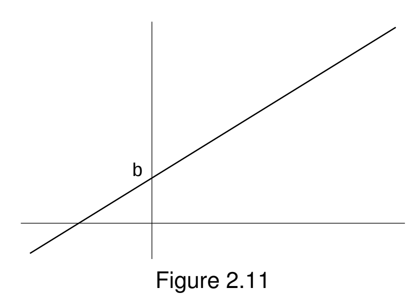


Figure 2.11

(3) **Quadratic Functions** A *quadratic function* is a function f given by

$$f(x) = ax^2 + bx + c,$$

where a, b and c are constants and $a \neq 0$.

The domain of every quadratic function is $\mathbb{R}$.

The range is $[k, \infty)$ if $a > 0$ and $(-\infty, k]$ if $a < 0$ where k is the y-coordinate of the vertex.

The graph is a parabola which opens upward if $a > 0$ and downward if $a < 0$.

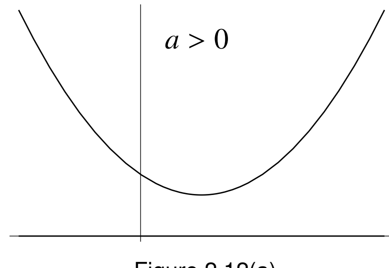


Figure 2.12(a)

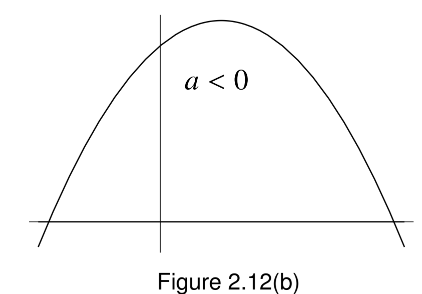


Figure 2.12(b)

Remark Besides using the completing square method to find the vertex, we can also use *differentiation* (see Chapter 5).

(4) **Polynomial Functions** A function f given by

$$f(x) = a_n x^n + a_{n-1} x^{n-1} + \cdots + a_1 x + a_0,$$

where $a_0, a_1, \ldots, a_n$ are constants with $a_n \neq 0$, is called a *polynomial function* of *degree n*.

If $n = 0$, f is a constant function.
If $n = 1$, f is a linear function.
If $n = 2$, f is a quadratic function.

Example Let $f(x) = x^3 - 3x^2 + x - 1$.

The graph of f is shown in Figure 2.13.

In Chapter 5, we will discuss how to sketch graphs of polynomial functions.

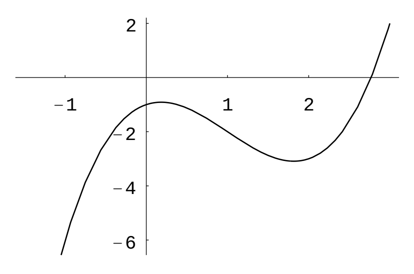


Figure 2.13

The domain of every polynomial function f is $\mathbb{R}$.

There are three possibilities for the range.

(a) If the degree is odd, then $\operatorname{ran}(f) = \mathbb{R}$.

(b) If the degree is even and positive, then

  (i) $\operatorname{ran}(f) = [k, \infty)$ if $a_n > 0$;

  (ii) $\operatorname{ran}(f) = (-\infty, k]$ if $a_n < 0$,

  where k is the y-coordinate of the lowest point for case (i), or the highest point for case (ii), of the graph.

Remark The constant function 0 is also considered to be a polynomial function. However, its degree is assigned to be $-\infty$ (for convenience of a rule for degree of product of polynomials).

(5) **Rational Functions** A *rational function* is a function f in the form

$$f(x) = \frac{p(x)}{q(x)},$$

where p and q are polynomial functions.

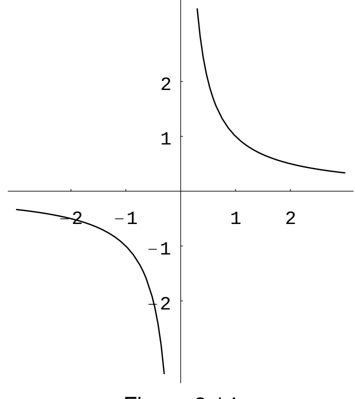


Figure 2.14

Example Let $f(x) = \frac{1}{x}$.

The domain of f is $\mathbb{R} \setminus \{0\}$.

The range of f is also $\mathbb{R} \setminus \{0\}$.

The graph consists of two curves, one in the first quadrant and the other in the third quadrant. It is symmetric about the origin. This is because $f(-x) = \frac{1}{-x} = -f(x)$.

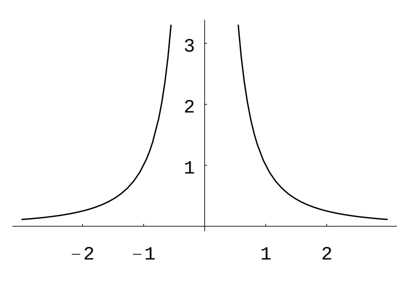


Figure 2.15

Example Let $f(x) = \frac{1}{x^2}$.

The domain of f is $\mathbb{R} \setminus \{0\}$.

The range of f is $(0, \infty)$.

The graph consists of two curves, one in the first quadrant and the other in the second quadrant. It is symmetric about the y-axis. This is because $f(-x) = \frac{1}{(-x)^2} = \frac{1}{x^2} = f(x)$.

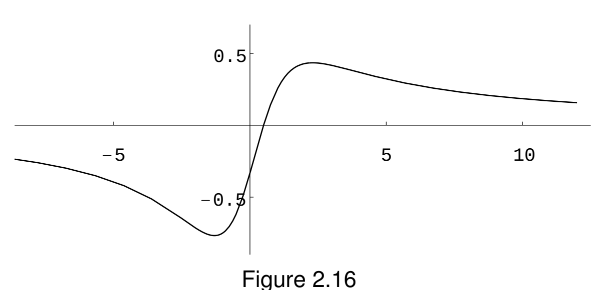


Figure 2.16

Example Let $f(x) = \frac{2x-1}{x^2+3}$.

The domain of f is $\mathbb{R}$.

The graph of f is shown in Figure 2.16. Note that when x is very large in magnitude, $f(x)$ is very small. This is because the degree of the numerator is smaller than that of the denominator. See Chapter 3 for more details.

The range of f can be found using the method described in Section 2.2. Alternatively, it can be found if the x-coordinates of the highest and lowest point are known. See Chapter 5 for more details.

(6) **Square-root Function**

Recall that for each positive real number x, there are two real numbers whose square is x. The two numbers are called the *square roots* of x. The *principle square root* of x, denoted by $\sqrt{x}$, is defined to be the positive square root of x.

Example The square roots of 9 are 3 and -3. The principle square root of 9 is 3, that is, $\sqrt{9} = 3$.

By convention, the principle square root of 0 is defined to be 0, that is, $\sqrt{0} = 0$.

The operation of taking principle square root can be considered as a function. Each nonnegative real number x can be used as an input and its corresponding output is $\sqrt{x}$.

Definition The *principle-square-root function*, denoted by sqrt, is the function given by

$$\text{sqrt}\,(x) = \sqrt{x}.$$

Remark

- Usually we use a single letter to denote a function. For functions that will be used very often, we create special notations for them. Usually, we use a few letters, taken from the names of the functions, to represent the functions.
- For simplicity, the principle-square-root function is also called the *square-root function*.
- Sometimes, the square-root function is also denoted by $\sqrt{x}$. Thus the notation $\sqrt{x}$ can have two different meanings:
  - ◊ a function (the square-root function)
  - ◊ a real number (the image of x under the square-root function)
- Sometimes, the square-root function is also denoted by $\sqrt{\cdot}$ (a dot inside $\sqrt{\ }$). Thus, $\sqrt{\cdot}(x) = \sqrt{x}$. The position of the dot indicates that the variable is put there.

The domain of the square-root function is $[0, \infty)$.

The range is also $[0, \infty)$.

The following steps describe how to draw the graph of $y = \sqrt{x}$.

(i) Square both sides to get $y^2 = x$.

The graph pf $y^2 = x$ is a parabola opening to the right. It can be obtained from the parabola given by $y = x^2$ by rotating 90° in the clockwise direction. Note that in the two equations, the role of x and y are interchanged.

(ii) The graph of $y = \sqrt{x}$ is the upper half of the parabola obtained in (i).

The lower part is not included because $\sqrt{x}$ is always non-negative. Squaring introduces extra points.

In the equation $y = \sqrt{x}$, it is implied implicitly that $x \geq 0$ and $y \geq 0$.

The graph of the equation is the following subset of $\mathbb{R}^2$:

$$\begin{aligned} \text{Graph} &= \{(x, y) \in \mathbb{R}^2 : y = \sqrt{x},\ x \geq 0,\ \ y \geq 0\} \\ &= \{(x, y) \in \mathbb{R}^2 : y^2 = x,\ x \geq 0,\ \ y \geq 0\} \end{aligned}$$

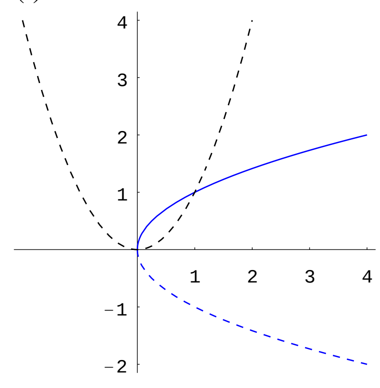


Figure 2.17

Example For each of the following equations, sketch its graph.

(a) $y = \sqrt{x} - 2$

(b) $y = \sqrt{x - 2}$

(c) $y = \sqrt{2 - x}$

Solution

(a) The graph is a half of a parabola. It is obtained by moving the graph of $y = \sqrt{x}$ two units down.

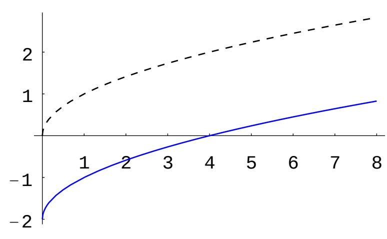


Figure 2.18

Remark Let a be a positive constant.

- The graph of $y = f(x) + a$ can be obtained from that of $y = f(x)$ by moving it a units up.
- The graph of $y = f(x) - a$ can be obtained from that of $y = f(x)$ by moving it a units down.

(b) Note that $\sqrt{x-2}$ is defined for $x \geq 2$ only. The graph of $y = \sqrt{x-2}$ is obtained by moving that of $y = \sqrt{x}$ two units to the right.

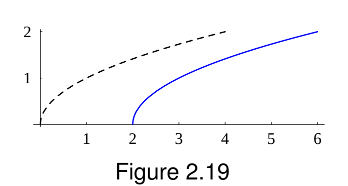


Figure 2.19

Remark Let a be a positive constant.

- The graph of $y = f(x - a)$ can be obtained from that of $y = f(x)$ by moving it a units to the right.
- The graph of $y = f(x + a)$ can be obtained from that of $y = f(x)$ by moving it a units to the left.

(c) Note that $\sqrt{2-x}$ is defined for $x \leq 2$ only. The graph of $y = \sqrt{2-x}$ and that of $y = \sqrt{x-2}$ are symmetric with respect to the vertical line $x = 2$.

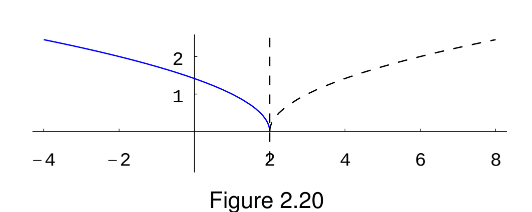


Figure 2.20

In general, two subsets of the plane are said to be *symmetric about a line* ℓ if for each point P belonging to any one of the two sets, there is a point Q belonging to the other set such that either $P = Q$ belongs to ℓ or

- the line segment PQ is perpendicular to ℓ;
- P and Q are equidistant from ℓ.

□

(7) **Exponential Functions** Let b be a positive real number different from 1. The *exponential function with base* b, denoted by $\exp_b$, is the function given by

$$\exp_b(x) = b^x.$$

The domain of every exponential function is $\mathbb{R}$.

The range of every exponential function is $(0, \infty)$.

The y-intercept of the graph of every exponential function is $(0, 1)$. This is because $b^0 = 1$.

Remark

- Because there are infinitely many exponential functions, one for each base, the notation $\exp_b$, where b is written as a subscript, indicates that the base is b. Thus, for example, $\exp_2$ and $\exp_3$ are the exponential functions with base 2 and 3 respectively.
- Sometimes, for convenience, we also write b^x to denote the exponential function with base b.

Example Consider $\exp_2$, the exponential function with base 2. It is the function from $\mathbb{R}$ to $\mathbb{R}$ given by $\exp_2(x) = 2^x$.

The graph of $\exp_2$ goes up (as x increases) and the rate that the graph goes up increases as x increases.

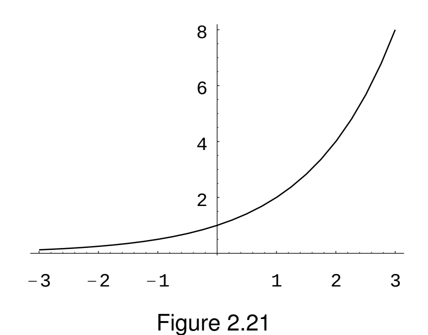


Figure 2.21

Example Consider $\exp_{\frac{1}{3}}$, the exponential function with base $\frac{1}{3}$. It is the function from $\mathbb{R}$ to $\mathbb{R}$ given by $\exp_{\frac{1}{3}}(x) = \left(\frac{1}{3}\right)^x$.

The graph of $\exp_{\frac{1}{3}}$ goes down (as x increases).

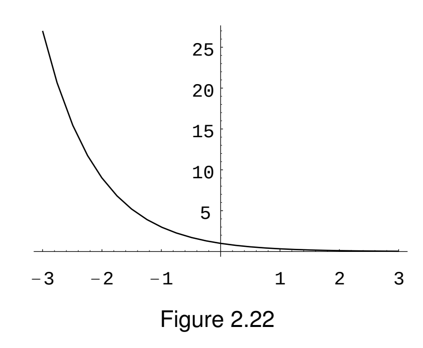


Figure 2.22

In Chapter 8, exponential functions will be discussed in more detail.

(8) **Logarithmic Function** Recall that for every positive real number x, there is a unique real number y such that $10^y = x$. Different positive x give different values of y. In this way, we obtain a function defined for all positive real numbers. This function, denoted by log, is called the *common logarithmic function*. For each positive real number x, $\log(x)$ is defined to be the unique real number such that $10^{\log(x)} = x$. That is, $\log(x) = y$ if and only if $y = 10^x$. For simplicity, $\log(x)$ is also written as $\log x$.

Remark Sometimes, for convenience, we also write $\log x$ to denote the common logarithmic function. So the notation $\log x$ has two different meanings. It can be a function (the log function) or a number (the image of x under the log function).

The domain of log is $(0, \infty)$.

The range is $\mathbb{R}$.

The graph of log is shown in Figure 2.23. As x increases, the graph goes up. The rate that the graph goes up decreases as x increases. Note that the x-intercept is $(1, 0)$. This is because $\log 1 = 0$.

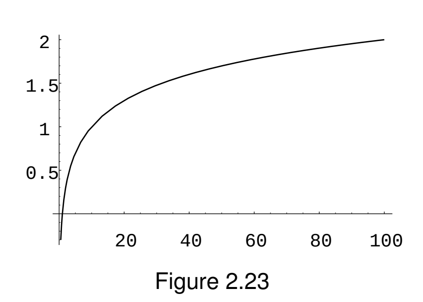


Figure 2.23

Remark In Chapter 8, logarithmic functions with bases other than 10 will be considered. Relation between exponential functions and logarithmic functions will be discussed.

(9) **Trigonometric Functions** Similar to the common logarithmic function, we use three letters sin, cos and tan to denote the sine, cosine and tangent functions respectively. Recall that $\sin(x)$ is defined to be the y-coordinate of the point on the unit circle $x^2 + y^2 = 1$ corresponding to the angle with measure x radians. More details on trigonometric functions can be found in Chapter 7.

For simplicity, we write $\sin(x) = \sin x$ etc.

- The sine function: sin

  The domain of the sine function is $\mathbb{R}$.

  The range is $[-1, 1]$.

  The graph of the sine function has a waveform as shown in Figure 2.24 (the symbol p stands for the number π). The graph is symmetric about the origin. This is because $\sin(-x) = -\sin x$. The graph crosses the x-axis infinitely often, at points with x-coordinates $0, \pm\pi, \pm 2\pi, \ldots$

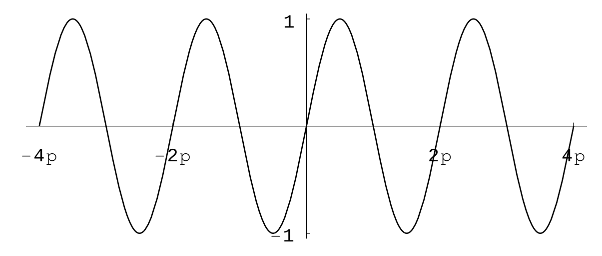


Figure 2.24

  The sine function is periodic with period 2π, that is, $\sin(x + 2\pi) = \sin x$ for all $x \in \mathbb{R}$.

  Definition If f is a function such that $f(x + p) = f(x)$ for all $x \in \text{dom}(f)$, where p is a positive constant, then we say that f is *periodic with period p*.

- The cosine function: cos

  The domain of the cosine function is $\mathbb{R}$.

  The range is $[-1, 1]$.

  The graph of the cosine function has a waveform. It is symmetric about the y-axis. This is because $\cos(-x) = \cos x$. The graph crosses the x-axis infinitely often, at points with x-coordinates $\pm\frac{\pi}{2}, \pm\frac{3\pi}{2}, \ldots$

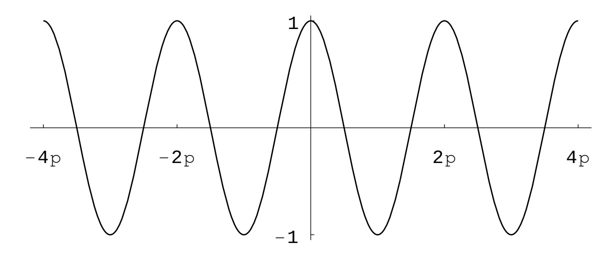


Figure 2.25

  The cosine function is periodic with period 2π, that is, $\cos(x + 2\pi) = \cos x$ for all $x \in \mathbb{R}$.

  Remark The graph of the cosine function can be obtained by shifting the graph of the sine function $\frac{\pi}{2}$ units to the left. This is because $\cos x = \sin\left(x + \frac{\pi}{2}\right)$ for all $x \in \mathbb{R}$.

(a) The tangent function: tan

Since $\tan x = \dfrac{\sin x}{\cos x}$, $\tan x$ is undefined at $x = \pm\dfrac{\pi}{2}, \pm\dfrac{3\pi}{2}, \ldots$

The domain of the tangent function is $\mathbb{R} \setminus \{\pm\frac{\pi}{2}, \pm\frac{3\pi}{2}, \ldots\}$.

The range is $\mathbb{R}$.

The tangent function is periodic with period π, that is, $\tan(x + \pi) = \tan x$ for all x belonging to the domain.

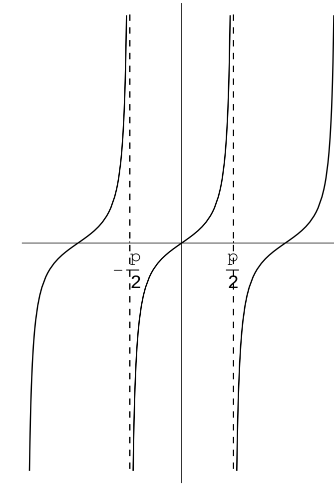


Figure 2.26

(10) **Absolute Value Function** The *absolute value function*, denoted by $|\cdot|$, is the function from $\mathbb{R}$ to $\mathbb{R}$ given by

$$|x| = \begin{cases} x & \text{if } x > 0, \\ 0 & \text{if } x = 0, \\ -x & \text{if } x < 0. \end{cases}$$

For each real number a, the number $|a|$ is called the *absolute value* of a.

In defining $|x|$, the domain $\mathbb{R}$ is divided into three disjoint subsets, namely $(0, \infty)$, $\{0\}$ and $(-\infty, 0)$.

(a) If $x \in (0, \infty)$ which means $x > 0$, then $|x|$ is defined to be x.

(b) If $x \in \{0\}$ which means $x = 0$, then $|x|$ is defined to be 0.

(c) If $x \in (-\infty, 0)$ which means $x < 0$, then $|x|$ is defined to be $-x$.

Functions defined in this way are called *piecewise-defined functions*.

Remark Unlike most functions, the image of a real number x under the absolute value function $|\cdot|$ is denoted by $|x|$. That is, $|x| = |\cdot|(x)$. Readers may compare this with the square-root function $\sqrt{\cdot}$, where $\sqrt{\cdot}(x) = \sqrt{x}$.

Example

(a) $|2| = 2$

(b) $|-3| = -(-3) = 3$

(c) $-|2| = -2$

(d) $-|-3| = -3$

(e) $|5 - (-7)| = |12| = 12$

(f) $\big||3 - 12| - |7 + 6|\big| = \big||-9| - |13|\big| = |9 - 13| = |-4| = 4$

Remark

(a) $|a|$ is always nonnegative.

(b) $|a| = |-a|$.

(c) $|a|$ is the distance from the point a to 0 on the real number line.

(d) $|a-b|$ is the distance between a and b.

(e) $\sqrt{a^2} = |a|$.

The domain of $|\cdot|$ is $\mathbb{R}$.

The range is $[0,\infty)$.

The graph of the absolute value function a V-shape figure. It is the union of the following three subsets of $\mathbb{R}^2$.

- $\{(x,y) \in \mathbb{R}^2 : x > 0 \text{ and } y = x\}$ which is the half-line in the first quadrant with slope equal to 1, starting from the origin but not including the origin.
- $\{(x,y) \in \mathbb{R}^2 : x = 0 \text{ and } y = 0\}$ which is the point $\{0,0\}$, that is, the origin.
- $\{(x,y) \in \mathbb{R}^2 : x < 0 \text{ and } y = -x\}$ which is the half-line in the second quadrant with slope equal to -1, starting from the origin but not including the origin.

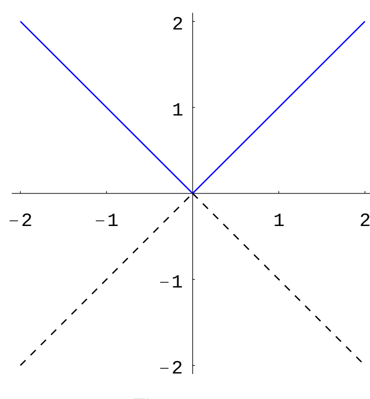


Figure 2.27

Remark We may also define the absolute value function in the following ways:

(i) $|x| = \begin{cases} x & \text{if } x \geq 0, \\ -x & \text{if } x < 0. \end{cases}$

(ii) $|x| = \begin{cases} x & \text{if } x > 0, \\ -x & \text{if } x \leq 0. \end{cases}$

(iii) $|x| = \begin{cases} x & \text{if } x \geq 0, \\ -x & \text{if } x \leq 0. \end{cases}$

In (i) or (ii), the domain $\mathbb{R}$ is divided into two disjoint subsets.

In (iii), although $\mathbb{R}$ is the union of $(-\infty,0]$ and $[0,\infty)$, the two subsets are not disjoint; the number 0 belongs to both. However, this will not cause any problem to define $|0|$ because if we use the first rule, we get $|0| = 0$ and if we use the second rule, we get $|0| = -0 = 0$. We say that $|0|$ is *well-defined* because its value does not depend on the choice of the rule.

Example For each of the following equations, sketch its graph.

(a) $y = 1 - |x|$

(b) $y = |x-1|$

Solution

(a) By the definition of the absolute value function, the equation is

$$y = \begin{cases} 1-x & \text{if } x > 0, \\ 1-0 & \text{if } x = 0, \\ 1-(-x) & \text{if } x < 0. \end{cases}$$

The graph is shown in Figure 2.28. It consists of the half-line $y = 1-x$, $x > 0$, the point $(0,1)$ and the half-line $y = 1+x$, $x < 0$.

Remark Alternatively, the graph can be obtained as follows:

- The graph of $y = -|x|$ and that of $y = |x|$ are symmetric about the x-axis. So the graph of $y = -|x|$ is an inverted V-shape figure.
- Move the inverted V-shape figure 1 unit up.

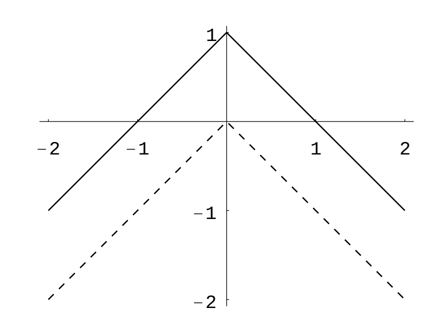
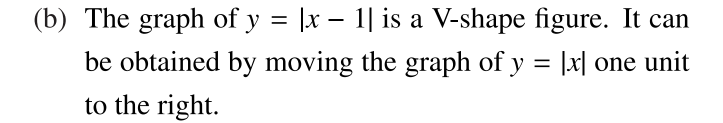


Figure 2.28



(b) The graph of $y = |x - 1|$ is a V-shape figure. It can be obtained by moving the graph of $y = |x|$ one unit to the right.

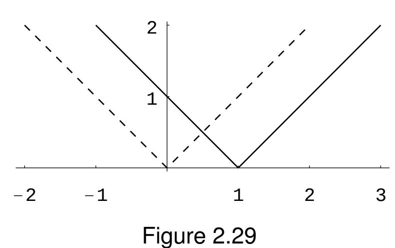


Figure 2.29

(11) **Piecewise-defined Functions** Below we give more examples of piecewise-defined functions.

Example Let $f : [-2, 6] \longrightarrow \mathbb{R}$ be the function given by

$$f(x) = \begin{cases} x^2 & \text{if } -2 \le x < 0 \\ 2x & \text{if } 0 \le x < 2 \\ 4 - x & \text{if } 2 \le x \le 6. \end{cases}$$

For each of the following, find its value:

(a) $f(-1)$

(b) $f(\frac{1}{2})$

(c) $f(3)$

Sketch the graph of f.

Solution

(a) $f(-1) = (-1)^2 = 1$

(b) $f(\frac{1}{2}) = 2 \cdot \frac{1}{2} = 1$

(c) $f(3) = 4 - 3 = 1$

The graph of f consists of three parts:

- the curve $y = x^2$, $-2 \le x < 0$ (part of a parabola);
- the line segment $y = 2x$, $0 \le x < 2$ (excluding the right endpoint);
- the line segment $y = 4 - x$, $2 \le x \le 6$.

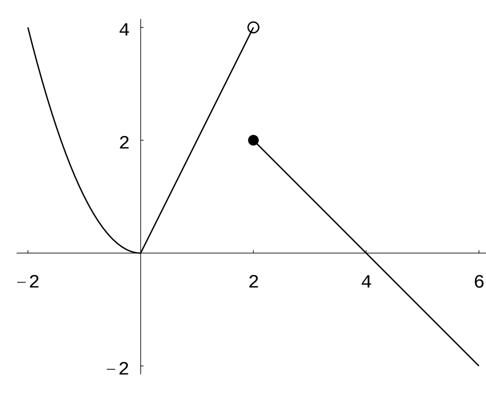


Figure 2.30

Remark In the figure, the little circle indicates that the point $(2, 4)$ is not included in the graph. The little dot (which can be omitted) emphasizes that the point $(2, 2)$ is included. □

The next example shows that piecewise-defined functions can be used in daily life. The piecewise-defined function in the example is called a *step function*. It jumps from one value to another.

Example Suppose the long-distance rate for a telephone call from City A to City B is \$1.4 for the first minute and \$0.9 for each additional minute or fraction thereof. If $y = f(t)$ is a function that indicates the total charge y for a call of t minutes' duration, sketch the graph of f for $0 < t \le 4\frac{1}{2}$.

Solution Note that

$$f(t) = \begin{cases} 1.4 & \text{if } 0 < t \le 1 \\ 2.3 & \text{if } 1 < t \le 2 \\ 3.2 & \text{if } 2 < t \le 3 \\ 4.1 & \text{if } 3 < t \le 4 \\ 5.0 & \text{if } 4 < t \le 4\frac{1}{2}. \end{cases}$$

The graph of $y = f(t)$ is shown in Figure 2.31.

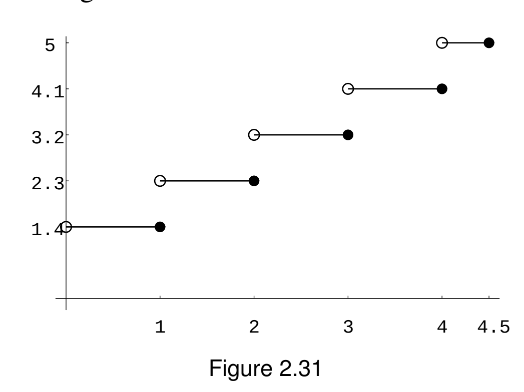


Figure 2.31

□

Remark The *ceiling* of a real number t, denoted by $\lceil t \rceil$, is defined to be the smallest integer greater than or equal to t. Using this notation, we have

$$f(t) = 1.4 + 0.9(\lceil t \rceil - 1).$$

Exercise 2.4

1. For each of the following equations, sketch its graph.
   - (a) $y = 2x - 3$
   - (b) $y + 3 = 2(x - 5)$
   - (c) $7x - 5y + 4 = 0$
   - (d) $y = |2x - 1| + 5$
   - (e) $x^2 = y^2$
   - (f) $y = -x^2$
   - (g) $y - 2 = -x^2$
   - (h) $y - 2 = -(x - 3)^2$
   - (i) $y = x^2 + 2x - 3$
   - (j) $y = \sqrt{1 - x^2}$
   - (k) $x = \sqrt{y}$
2. For each of the following equations, use a computer software to sketch its graph.
   - (a) $y = x^3$
   - (b) $y = x^3 - 2x^2 - 3x + 4$
   - (c) $y = 2x^3 - x + 5$
   - (d) $y = x^3 - 3x^2 + 3x - 1$
   - (e) $y = -x^3$
   - (f) $y = -x^3 + 2x^2 + 3x - 4$
   - (g) $y = x^4$
   - (h) $y = x^4 - x^3 - x^2 + x + 1$
   - (i) $y = x^4 - 3x^3 + 2x^2 + x - 1$
   - (j) $y = x^4 - 4x^3 + 6x^2 - 4x$
   - (k) $y = -x^4$
   - (l) $y = -x^4 - 2x^3 + 3x$

Can you generalize the results for graphs of polynomial functions of degree $3, 4, \ldots$ *?*

3. Let $f(x) = \frac{2x-1}{x^2+3}$. The graph of f is shown on page 55. Note that there is a highest point and a lowest point. Find the coordinates of these two points. *Hint: consider the range of f*
   The points are called *relative extremum points*. An easy way to find their coordinates is to use *differentiation*, see Chapter 5.
4. An object is thrown upward and its height $h(t)$ in meters after t seconds is given by $h(t) = 1 + 4t - 5t^2$.
   (a) When will the object hit the ground?
   (b) Find the maximum height attained by the object.
5. The manager of an 80-unit apartment complex is trying to decide what rent to charge. Experience has shown that at a rent of \$20000, all the units will be full. On the average, one additional unit will remain vacant for each \$500 increase in rent.
   (a) Let n represent the number of \$500 increases.
   Find an expression for the total revenue R from all the rented apartments.
   What is the domain of R?
   (b) What value of n leads to maximum revenue?
   What is the maximum revenue?

2.5 Compositions of Functions

Consider the function f given by

$$f(x) = \sin^2 x.$$

Recall that $\sin^2 x = (\sin x)^2$. For each input x, to find the output $y = f(x)$,

(1) first calculate $\sin x$, call the resulted value u;

(2) and then calculate u^2.

These two steps correspond to two functions:

(1) $u = \sin x$;

(2) $y = u^2$.

Given two functions, we can “combine” them by letting one function acting on the output of the other.

Definition Let f and g be functions such that the codomain of f is a subset of the domain of g. The *composition* of g with f, denoted by $g \circ f$, is the function given by

$$(g \circ f)(x) = g(f(x)). \tag{2.5.1}$$

The right-side of (2.5.1) is read “*g of f of x*”.

Figure 2.32 indicates that f is a function from A to B and g is a function from C to D where $B \subseteq C$.

For each element x of A (the domain of f), its image $f(x)$ is an element of B. Because $B \subseteq C$, $f(x)$ is an element of C which is the domain of g. Therefore, $f(x)$ can be used an input for the function g and the output $g(f(x))$ is an element of D. Thus $g \circ f$ is a function from A to D.

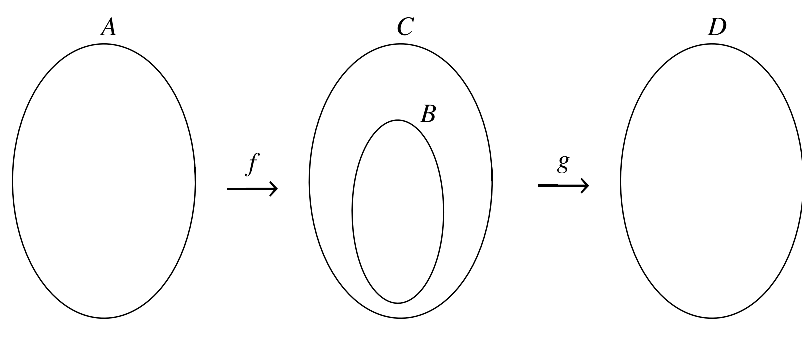


Figure 2.32

Remark In the above definition, the condition that *codomain of* $f \subseteq$ *domain of* g can be relaxed. In order to consider $g(f(x))$, we only need $f(x)$ *belong to the domain of* g. This is satisfied if $\operatorname{ran}(f) \subseteq \operatorname{dom}(g)$.

In the following two examples, the domains of both f and g are equal to $\mathbb{R}$. Therefore we can consider $g \circ f$ as well as $f \circ g$. The first example illustrates how the functions $\sin x$ and x^2 are used as building blocks for the more complicated function $\sin^2 x$ (see the discussion preceding the above definition). The second example shows that composition of functions is not commutative, that is, $f \circ g \neq g \circ f$ in general.

Example Let $f : \mathbb{R} \longrightarrow \mathbb{R}$ and $g : \mathbb{R} \longrightarrow \mathbb{R}$ be given by

$$f(x) = \sin x \qquad \text{and} \qquad g(x) = x^2.$$

Then we have
$$\begin{aligned}(g \circ f)(x) &= g(f(x)) \\ &= g(\sin x) \\ &= (\sin x)^2 = \sin^2 x.\end{aligned}$$

Example Let $f(x) = x^2$ and $g(x) = 2x + 1$. Find $(f \circ g)(x)$ and $(g \circ f)(x)$.

Solution By the definition of composition, we have

$$\begin{aligned}(f \circ g)(x) &= f(g(x)) = f(2x+1) \\ &= (2x+1)^2 = 4x^2 + 4x + 1 \\ (g \circ f)(x) &= g(f(x)) = g(x^2) \\ &= 2x^2 + 1.\end{aligned}$$

□

Remark If the range of f is not contained in the domain of g, then we have to restrict f to a smaller set so that for every x in that set, $f(x)$ belongs to the domain of g. The domain of $g \circ f$ is taken to be the following:

$$\operatorname{dom}(g \circ f) = \{x \in \operatorname{dom}(f) : f(x) \in \operatorname{dom}(g)\}.$$

Example Let $f(x) = x + 1$ and $g(x) = \sqrt{x}$. Find the domain of $g \circ f$.

Solution Note that the domain of f is $\mathbb{R}$ and the domain of g is $[0, \infty)$. Thus the domain of $g \circ f$ is

$$\begin{aligned}\operatorname{dom}(g \circ f) &= \{x \in \mathbb{R} : x + 1 \in [0, \infty)\} \\ &= \{x \in \mathbb{R} : x \geq -1\} \\ &= [-1, \infty).\end{aligned}$$

□

Remark Alternatively, note that $(g \circ f)(x) = \sqrt{x+1}$. Thus we have $\operatorname{dom}(g \circ f) = \{x \in \mathbb{R} : x + 1 \geq 0\}$.

Exercise 2.5

1. Let $f(x) = x^2 + 1$ and $g(x) = x + 1$. Find the following:
   (a) $(f \circ g)(1)$ (b) $(g \circ f)(1)$
   (c) $(f \circ g)(x)$ (d) $(g \circ f)(x)$
   (e) $(f \circ g)(a^2)$ (f) $(g \circ f)(\sqrt{a})$
2. For each of the following, find $f(x)$ and $g(x)$ where $g(x)$ is in the form x^r with $r \neq 1$ such that $(g \circ f)(x)$ equals the given expression.
   (a) $\sqrt{x^2+1}$ (b) $\dfrac{1}{x+1}$

2.6 Inverse Functions

Let f be the function given by $f(x) = 2^x$ for $x \in \mathbb{R}$. Consider the equation $y = f(x)$, that is,

$$y = 2^x. \tag{2.6.1}$$

- In Equation (2.6.1), if we put $x = x_1$, we obtain the corresponding value of y, namely $y_1 = 2^{x_1}$. The number x_1 is an input of the function f and the value y_1 is the corresponding output.
- Now we consider the reverse problem. If we put $y = y_1$, can we find a real number x_1 such that $2^{x_1} = y_1$?
  (1) If $y_1 \leq 0$, there is no solution because 2^x is always positive.
  (2) If $y_1 > 0$, there is exactly one solution because f is injective (see definition below) and its range is $(0, \infty)$.

Definition Let f be a function. We say that f is *injective* if the following condition is satisfied:

$(*)$ $x_1, x_2 \in \operatorname{dom}(f)$ and $x_1 \neq x_2 \Longrightarrow f(x_1) \neq f(x_2)$.

Condition $(*)$ means that different elements of the domain are mapped to different elements of the codomain. It is equivalent to the following condition:

$(**)$ $x_1, x_2 \in \operatorname{dom}(f)$ and $f(x_1) = f(x_2) \Longrightarrow x_1 = x_2$.

Example Let $f(x) = 2^x$. The domain of f is $\mathbb{R}$. The function f is injective. This is because if $x_1, x_2 \in \mathbb{R}$ and $x_1 \neq x_2$, then $2^{x_1} \neq 2^{x_2}$.

Example Let $g(x) = x^2$. The domain of g is $\mathbb{R}$. The function g is not injective. This is because $-1 \neq 1$ (both are elements of $\mathbb{R}$), but $g(-1) = g(1)$.

To show that a function f is injective, we have to consider all x_1, x_2 belonging to the domain with $x_1 \neq x_2$ and check that $f(x_1) \neq f(x_2)$. However, to show that a function g is not injective, it suffices to find two different elements x_1, x_2 of the domain such that $g(x_1) = g(x_2)$. Below we give a geometric method to determine whether a function is injective or not.

Horizontal Line Test Let $f : X \longrightarrow \mathbb{R}$ be a function where $X \subseteq \mathbb{R}$. Then f is injective if and only if every horizontal line intersects the graph of f in at most one point.

Example The following two figures show the graphs of f and g in the last two examples. It is easy to see from the Horizontal Line Test that f is injective whereas g is not injective.

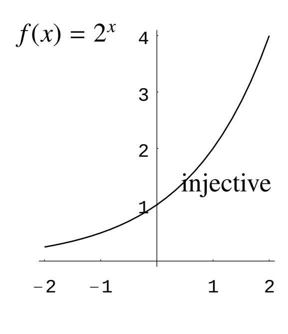


Figure 2.33

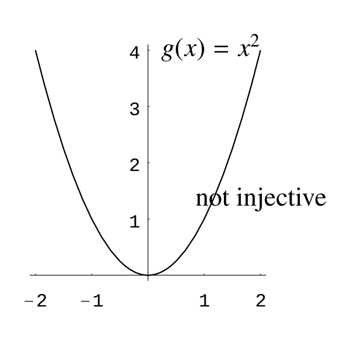


Figure 2.34

Let f be an injective function. Then given any element y of $\text{ran}(f)$, there is exactly one element x of $\text{dom}(f)$ such that $f(x) = y$. This means that if we use an element y of $\text{ran}(f)$ as input, we get one and only one output x. The function obtained in this way is called the inverse of f.

Definition Let $f : X \longrightarrow Y$ be an injective function and let Y_1 be the range of f. The *inverse (function)* of f, denoted by f^{-1}, is the function from Y_1 to X such that for every $y \in Y_1$, $f^{-1}(y)$ is the unique element of X satisfying $f(f^{-1}(y)) = y$.

The following figure indicates a function f from a set X to a set Y. Assuming that f is injective, for each y belonging to the range of f, there is one and only one element x of X such that $f(x) = y$. This element x is defined to be $f^{-1}(y)$. That is,

$$f^{-1}(y) = x \quad \text{if and only if} \quad f(x) = y.$$

Remark

(1) For every $x \in X$, we have $(f^{-1} \circ f)(x) = x$.
For every $y \in Y_1$, we have $(f \circ f^{-1})(y) = y$.

(2) f^{-1} is injective and $(f^{-1})^{-1}(x) = f(x)$ for all $x \in \text{dom}(f)$.

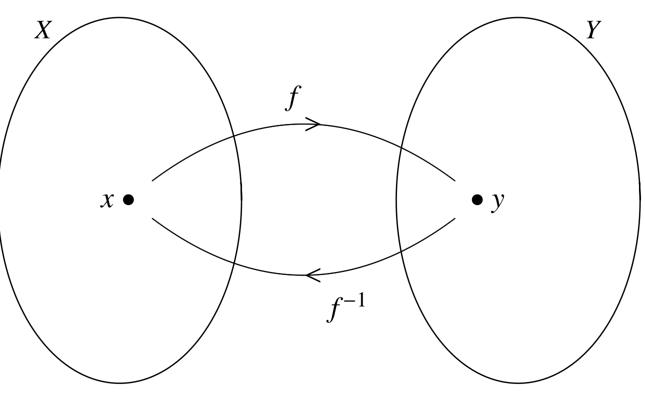


Figure 2.35

Steps to find inverse functions

Let $f : X \longrightarrow \mathbb{R}$ be an injective function where $X \subseteq \mathbb{R}$. To find the inverse function of f means to find the domain of f^{-1} as well as a formula for $f^{-1}(y)$. If the formula for $f(x)$ is not very complicated, $\text{dom}(f^{-1})$ and $f^{-1}(y)$ can be found by solving the equation $y = f(x)$ for x.

(*Step 1*) Put $y = f(x)$.

(*Step 2*) Solve x in terms of y. The result will be in the form x = an expression in y.

(*Step 3*) From the expression in y obtained in Step 2, the range of f can be determined. This is the domain of f^{-1}. The required formula is $f^{-1}(y)$ = the expression in y obtained in Step 2.

Remark Steps 1 and 2 can be used to find range of a function. If the function is not injective, the expression in y obtained in Step 2 does not give a function; some y give more than one values of x.

Example Let $f(x) = 2x^3 + 1$. Find the inverse of f.

Solution The domain of f is $\mathbb{R}$. It is not difficult to show that f is injective and that the range of f is $\mathbb{R}$. These two facts can also be seen from the following steps:

Put $y = f(x)$. That is, $y = 2x^3 + 1$.

Solve for x:
$$\begin{aligned} y-1 &= 2x^3 \\ \frac{y-1}{2} &= x^3 \\ \sqrt[3]{\frac{y-1}{2}} &= x \quad (x \text{ can be solved for all real numbers } y) \end{aligned}$$

Thus we have $\operatorname{dom}(f^{-1}) = \mathbb{R}$ and $f^{-1}(y) = \sqrt[3]{\dfrac{y-1}{2}}$. □

Example Let $g : [0, \infty) \longrightarrow \mathbb{R}$ be the function given by $g(x) = x^2$. Find the inverse of g.

Solution Because the domain of g is $[0, \infty)$, the function g is injective . Moreover, the range of g is $[0, \infty)$. These two facts can also be seen from the following steps:

Put $y = g(x)$. That is, $y = x^2$. Note that $y \geq 0$ and that $x \geq 0$ since $x \in \operatorname{dom}(f)$.

Solve for x:
$$\begin{aligned} y &= x^2, \quad y \geq 0,\ x \geq 0 \\ \sqrt{y} &= x \quad (x \text{ can be solved if and only if } y \geq 0,\ x = -\sqrt{y} \text{ is rejected}) \end{aligned}$$

Thus we have $\operatorname{dom}(g^{-1}) = [0, \infty)$ and $g^{-1}(y) = \sqrt{y}$. □

Remark Usually, we use x to denote the independent variable of a function. For the above examples, we may write $f^{-1}(x) = \sqrt[3]{\frac{x-1}{2}}$ and $g^{-1}(x) = \sqrt{x}$.

Caution $f^{-1}(x) \neq \dfrac{1}{f(x)}$

Remark We use $\sin^{-1}$ or arcsin to denote the inverse of sin etc. Although the sine function is not injective, we can make it injective by restricting the domain to $[-\frac{\pi}{2}, \frac{\pi}{2}]$.

$x = \sin^{-1} y$ means $\sin x = y$ and $-\frac{\pi}{2} \leq x \leq \frac{\pi}{2}$. The domain of $\sin^{-1}$ is $[-1, 1]$ because $-1 \leq \sin x \leq 1$.

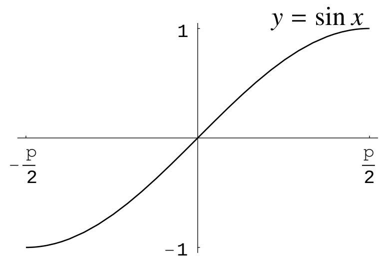


Figure 2.36

FAQ Why do we use the notation f^{-1}?

Answer The following example gives a reason why we use such a notation. Let $f(x) = 2x$. Then f is injective and its inverse is given by $f^{-1}(x) = \frac{1}{2}x$. The multiplicative constant $\frac{1}{2}$ is 2^{-1}.

Another reason is to have the "index law" (details omitted):

$$f^m \circ f^n = f^{m+n} \qquad \text{for } m, n \in \mathbb{Z}.$$

□

Graph of the inverse function

Let $f : X \longrightarrow \mathbb{R}$ be a function, where $X \subseteq \mathbb{R}$. Then its graph is a subset of the plane. If, in addition, f is injective, then f has an inverse and $\operatorname{dom}(f^{-1}) \subseteq \mathbb{R}$. Hence the graph of f^{-1} is also a subset of the plane. There is a nice relationship between the graph of f and that of f^{-1}:

(∗) The graph of f and the graph of f^{-1} are symmetric about the line $x = y$.

Reason Suppose $P(a, b)$ belongs to the graph of f. This means that $b = f(a)$ or equivalently, $a = f^{-1}(b)$. Thus $Q(b, a)$ belongs to the graph of f^{-1}. It is straightforward to show that the line segment PQ is perpendicular to the line $y = x$ (denoted by ℓ) and that P and Q are equidistant from ℓ. □

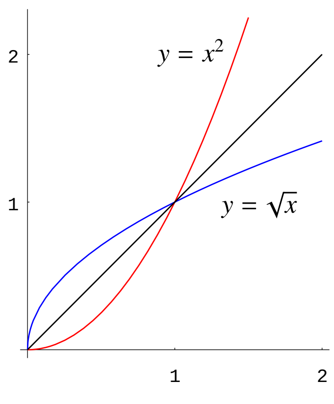


Figure 2.37

Figure 2.37 is an illustration for (∗). The function $f : [0, \infty) \longrightarrow \mathbb{R}$ given by $f(x) = x^2$ is injective. Its range is $[0, \infty)$. The domain of f^{-1} is $[0, \infty)$ and $f^{-1}(x) = \sqrt{x}$.

Exercise 2.6

1. For each of the following functions f, determine whether it is injective or not.

   (a) $f(x) = x^3 + 2x$ (b) $f(x) = x^2 - 5$

2. For each of the following functions f, find its inverse.

   (a) $f(x) = 3x - 2$ (b) $f(x) = x^5 + 3$

   (c) $f(x) = 1 + 2x^{\frac{1}{7}}$ (d) $f(x) = \sqrt[3]{2x^3 - 1}$

2.7 More on Solving Equations

In this section, we will consider *fractional equations* and *radical equations*. In solving equations, if there is a one-sided implication ($\Longrightarrow$) in any one of the steps, we have to check solution. If all the steps are two-sided implications ($\Longleftrightarrow$), there is no need to check solution.

Example For each of the following equations, find its solution set.

(1) $\dfrac{5}{x-2} = \dfrac{10}{x+3}$

(2) $\dfrac{x}{x-1} + \dfrac{2}{x} = \dfrac{1}{x^2 - x}$

Solution

(1) Multiplying both sides of the given equation by $(x-2)(x+3)$, we get

$$\begin{aligned}
\frac{5}{x-2} &= \frac{10}{x+3} \\
\frac{5}{x-2}\cdot(x-2)(x+3) &= \frac{10}{x+3}\cdot(x-2)(x+3) \\
5(x+3) &= 10(x-2) \\
5x+15 &= 10x-20 \\
35 &= 5x \\
x &= 7
\end{aligned}$$

By direct substitution, we see that 7 is the solution to the given equation.
The solution set is $\{7\}$.

(2) Multiplying both sides by $x(x-1)$ which is the *LCM* of the denominators of the terms appearing in the equation, we get

$$\begin{aligned}
\frac{x}{x-1}+\frac{2}{x} &= \frac{1}{x^2-x} \\
\left(\frac{x}{x-1}+\frac{2}{x}\right)\cdot x(x-1) &= \frac{1}{x^2-x}\cdot x(x-1) \\
x^2+2(x-1) &= 1 \qquad (2.7.1) \\
x^2+2x-3 &= 0 \\
(x-1)(x+3) &= 0 \\
x=1 \quad \text{or} &\quad x=-3
\end{aligned}$$

By direct substitution, we see that -3 is a solution but 1 is not a solution to the given equation.
The solution set is $\{-3\}$. □

FAQ Why do we need to check solution?

Answer When we multiply both sides by $x(x-1)$, extra solutions may be introduced. Solutions to (2.7.1) may not be solutions to the original equation. This is because

$$a=b \Longrightarrow ac=bc, \quad \text{but the converse is true only if } c\neq 0.$$

□

FAQ Can we add some conditions on x so that the implication can go backward?

Answer In the given equation, it is understood that $x\neq 0$ and $x\neq 1$. This is because the domain of both $f(x)=\frac{x}{x-1}+\frac{2}{x}$ (the left-side) and $g(x)=\frac{1}{x^2-x}$ (the right-side) are $\mathbb{R}\setminus\{0,1\}$. Adding the conditions $x\neq 0$ and $x\neq 1$, each step below is a two-sided implication:

$$\begin{aligned}
\frac{x}{x-1}+\frac{2}{x}=\frac{1}{x^2-x} &\iff x^2+2(x-1)=1 \quad \text{and} \quad x\neq 0 \quad \text{and} \quad x\neq 1 \\
&\ \ \vdots \\
&\iff (x-1)(x+3)=0 \quad \text{and} \quad x\neq 0 \quad \text{and} \quad x\neq 1 \\
&\iff x=-3.
\end{aligned}$$

□

Example For each of the following equations, find its solution set.

(1) $\sqrt{x^2 - 7} + x = 7$

(2) $\sqrt{x} - \sqrt{x - 3} = 3$

Solution

(1) Rearranging terms and squaring both sides, we get

$$\begin{aligned} \sqrt{x^2 - 7} + x &= 7 \\ \sqrt{x^2 - 7} &= 7 - x \\ x^2 - 7 &= (7 - x)^2 \\ x^2 - 7 &= 49 - 14x + x^2 \\ -56 &= -14x \\ x &= 4 \end{aligned}$$

By direct substitution, we see that 4 is the solution to the given equation.
The solution set is $\{4\}$.

(2) Rearranging terms and squaring both sides, we get

$$\begin{aligned} \sqrt{x} - \sqrt{x - 3} &= 3 \\ \sqrt{x} - 3 &= \sqrt{x - 3} && (2.7.2) \\ \left(\sqrt{x} - 3\right)^2 &= x - 3 && (2.7.3) \\ x - 6\sqrt{x} + 9 &= x - 3 \\ -6\sqrt{x} &= -12 \\ \sqrt{x} &= 2 \\ x &= 4 \end{aligned}$$

By direct substitution, we see that 4 is not a solution to the given equation.
The equation has no solution. The solution set is $\emptyset$. □

FAQ Why do we need to check solution?

Answer When we square both sides of an equation, extra solutions may be introduced. Solutions to (2.7.3) may not be solutions to (2.7.2). This is because

$$a = b \Longrightarrow a^2 = b^2, \quad \text{but the converse is true only if } a \text{ and } b \text{ have the same sign.}$$

□

FAQ Can we add some conditions so that the implication can go backward?

Answer In the equation, it is understood that $x \geq 0$ and $x - 3 \geq 0$. Moreover, it is also understood that $\sqrt{x} \geq 3$. This can be seen easily from (2.7.2) Simplifying the three conditions: $x \geq 0$, $x \geq 3$ and $x \geq 9$, we get $x \geq 9$.

Adding this condition, each step below is a two-sided implication:

$$\begin{aligned}
\sqrt{x} - \sqrt{x-3} = 3 &\iff \sqrt{x} - 3 = \sqrt{x-3} \\
&\iff (\sqrt{x} - 3)^2 = x - 3 \quad \text{and} \quad x \geq 9 \\
&\iff x - 6\sqrt{x} + 9 = x - 3 \quad \text{and} \quad x \geq 9 \\
&\ \ \vdots \\
&\iff \sqrt{x} = 2 \quad \text{and} \quad x \geq 9 \\
&\iff x = 4 \quad \text{and} \quad x \geq 9
\end{aligned}$$

From this we see that there is no solution. □

Exercise 2.7

1. For each of the following equations, find its solution set.
   (a) $(2x+1)(x-2) = x(x+2)$
   (b) $(2x+1)(x-2) = x(x-2)$
   (c) $\frac{1}{x+1} = \frac{2}{x+2}$
   (d) $\frac{x}{x+2} - \frac{x}{x-2} = \frac{-4x}{x^2-4}$
   (e) $3 - \sqrt{2x+5} = 0$
   (f) $\sqrt{x^2-9} + x = 9$
   (g) $\sqrt{x^2-9} + 9 = x$
   (h) $\sqrt{x+5} + 1 = 2\sqrt{x}$
   (i) $x^6 - 9x^3 + 8 = 0$
2. Let $C(q) = 2q + 12$ be the cost to produce q units of a product and let $R(q) = 10q - q^2$ be the revenue.
   (a) Find the profit (function).
   (b) Find the break-even quantity.
3. The stopping distance y in feet of a car traveling at x mph is described by the equation

$$y = 0.056057x^2 + 1.06657x.$$

   (a) Find the stopping distance for a car traveling at 35 mph.
   (b) How fast can one drive if one needs to be certain of stopping within 200 ft?
4. Find the right-angle triangle such that the sides adjacent to the right angle differ by 1 unit and the perimeter is 12 units.

Chapter 3

Limits

Calculus is the study of *differentiation* and *integration* (this is indicated by the Chinese translation of "calculus"). Both concepts of differentiation and integration are based on the idea of *limit*. In this chapter, we use an intuitive approach to consider limits, omitting the more difficult ϵ-δ definition.

3.1 Introduction

In this section, we introduce the idea of limit by considering two problems. The first problem is to "find" the velocity of an object at a particular instant. The idea is related to *differentiation*. The second problem is to "find" the area under the graph of a curve (and above the x-axis). The idea is related to *integration*.

Problem 1 Suppose an object moves along the x-axis and its *displacement* (in meters) s at time t (in seconds) is given by

$$s(t) = t^2, \qquad t \geq 0.$$

We want to consider its *velocity* at a certain time instant, say at $t = 2$.

Idea Velocity (or speed) is defined by

$$\text{velocity} = \frac{\text{distance traveled}}{\text{time elapsed}} \tag{3.1.1}$$

(3.1.1) can only be applied to find average velocities over time intervals. We (still) don't have a definition for *velocity at* $t = 2$.

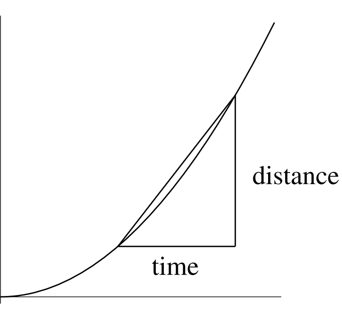


Figure 3.1

To define the velocity at $t = 2$, we consider short time intervals about $t = 2$, say from $t = 2$ to $t = 2 + \frac{1}{2^n}$. Using (3.1.1), we can compute the average velocity over the time intervals $[2, 2.5]$, $[2, 2.25]$ etc.

n	Time interval	Velocity
1	[2,2.5]	4.5 m/s
2	[2,2.25]	4.25 m/s
3	[2,2.125]	4.125 m/s
4	[2,2.0625]	4.0625 m/s
⋮		

In general, the velocity v_n over the time interval $[2, 2+\frac{1}{2^n}]$ is

$$\begin{aligned} v_n &= \frac{(2+\frac{1}{2^n})^2 - 2^2}{\frac{1}{2^n}} \\ &= \frac{4 + 2\cdot 2\cdot \frac{1}{2^n} + \left(\frac{1}{2^n}\right)^2 - 4}{\frac{1}{2^n}} = 4 + \frac{1}{2^n}. \end{aligned}$$

It is clear that if n is very large (that is, if the time interval is very short), v_n is very close to 4. The velocity, called the *instantaneous velocity*, at $t = 2$ is (defined to be) 4.

Problem 2 Find the area of the region that lies under the curve $y = x^2$ and above the x-axis for x between 0 and 1.

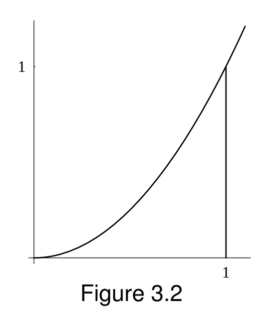


Figure 3.2

Idea Similar to the idea in Problem 1, we use approximation to find/define area. First we divide the interval [0, 1] into finitely many subintervals of equal lengths:

$$\left[0, \frac{1}{n}\right], \quad \left[\frac{1}{n}, \frac{2}{n}\right], \quad \left[\frac{2}{n}, \frac{3}{n}\right], \quad \ldots, \quad \left[\frac{n-1}{n}, 1\right].$$

For each subinterval $\left[\frac{i-1}{n}, \frac{i}{n}\right]$, we consider the rectangular region with base on the subinterval and height $\left(\frac{i-1}{n}\right)^2$ (the largest region that lies under the curve). If we add the area of these rectangular regions, the sum is smaller than that of the required region. However, if n is very large, the error is very small and we get a good approximation for the required area.

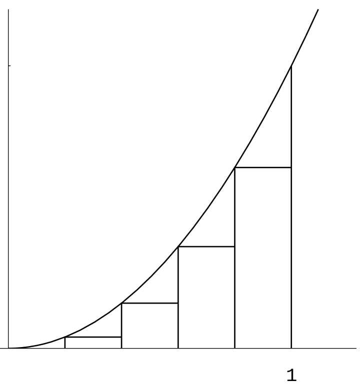


Figure 3.3

The following table gives the sum S_n of the areas of the rectangular regions (correct to 3 decimal places) for several values of n.

n	Sum of areas
2	0.125
3	0.185
4	0.219
⋮	
10	0.285
⋮	
100	0.328
⋮	
500	0.332

In general, if there are n subintervals, the sum S_n is

$$\begin{aligned} S_n &= \frac{1}{n}\cdot 0^2 + \frac{1}{n}\cdot\left(\frac{1}{n}\right)^2 + \frac{1}{n}\cdot\left(\frac{2}{n}\right)^2 + \cdots + \frac{1}{n}\cdot\left(\frac{n-1}{n}\right)^2 \\ &= \frac{1^2 + 2^2 + \cdots + (n-1)^2}{n^3} \\ &= \frac{n(n-1)(2n-1)}{6n^3} \qquad \text{By Sum of Squares Formula} \\ &= \frac{2n^3 - 3n^2 + n}{6n^3} \\ &= \frac{1}{3} - \frac{1}{2n} + \frac{1}{6n^2} \end{aligned}$$

It is clear that if n is very large (so that the error is small), S_n is very close to $\frac{1}{3}$.

Conclusion The area of the region is $\frac{1}{3}$.

Sum of Squares Formula $1^2 + 2^2 + \cdots + n^2 = \dfrac{n(n+1)(2n+1)}{6}$

Exercise 3.1

*1. In the first problem: If we consider short time intervals other than $[2, 2 + \frac{1}{2^n}]$, for example consider $[2, 2 + \frac{1}{n}]$ or $[2 - \frac{1}{n^2}, 2]$, do we get the same result?

*2. In the second problem:

(a) For each subinterval, we use the value of y at the left endpoint as the height of the rectangular region. If we take the right endpoint instead (we will get a surplus in this case), do we get the same result? How about taking an arbitrary point in each subinterval?

(b) In finding the approximations, we divide $[0, 1]$ into equal subintervals. How about dividing it into unequal subintervals?

3.2 Limits of Sequences

In the last section, we obtained formulas for v_n and S_n in Problems 1 and 2 respectively. Each of these formulas gives a *sequence* (which is a special type of function). To consider the behavior of a sequence for large n, we introduce the concept of *limit of a sequence*.

Definition

- A *sequence* is a function whose domain is $\mathbb{Z}_+$ (the set of all positive integers).
- A *sequence of real numbers* is a sequence whose codomain is $\mathbb{R}$.

A sequence of real numbers is a function from $\mathbb{Z}_+$ to $\mathbb{R}$. In this course, we will not consider sequences with codomains different from $\mathbb{R}$. Thus, in what follows, a sequence means a sequence of real numbers.

Let $f : \mathbb{Z}_+ \longrightarrow \mathbb{R}$ be a sequence. For each positive integer n, the value $f(n)$ is called the *nth term* of the sequence and is usually denoted by a small letter together with n in the subscript, for example a_n. The sequence is also denoted by $(a_n)_{n=1}^{\infty}$ because if we know all the a_n's, then we know the sequence.

Sometimes, we represent a sequence $(a_n)_{n=1}^{\infty}$ by listing a few terms in the sequence:

$$a_1,\ a_2,\ a_3,\ a_4,\ a_5,\ \ldots$$

In the following example, the sequences $(a_n)_{n=1}^{\infty}$ and $(b_n)_{n=1}^{\infty}$ are the sequences obtained in Problem 1 and Problem 2 in the last section respectively.

Example

(1) Let $a_n = 4 + \frac{1}{2^n}$. The sequence $(a_n)_{n=1}^{\infty}$ can be represented by

$$\frac{9}{2},\ \frac{17}{4},\ \frac{33}{8},\ \frac{65}{16},\ \ldots \tag{3.2.1}$$

(2) Let $b_n = \frac{1}{3} - \frac{1}{2n} + \frac{1}{6n^2}$. The sequence $(b_n)_{n=1}^{\infty}$ can be represented by

$$0,\ \frac{1}{8},\ \frac{5}{27},\ \frac{7}{32},\ \frac{6}{25},\ \ldots \tag{3.2.2}$$

Remark

- It is not a good way to describe a sequence by listing a few terms in the sequence. For example, in (3.2.1) or (3.2.2), it may not be easy to find a formula for the nth term. Moreover, different people may obtain different formulas. It is better to describe a sequence by writing down a formula for the nth term explicitly.
- To denote a sequence, some authors use the notation $\{a_n\}_{n=1}^{\infty}$ instead of $(a_n)_{n=1}^{\infty}$.

Definition A sequence $(a_n)_{n=1}^{\infty}$ is said to be *convergent* if there exists a real number L such that

($*$) a_n is arbitrarily close to L if n is sufficiently large.

Remark Condition ($*$) means that we can make $|a_n - L|$ as small as we want by taking n large enough. For the sequence $(a_n)_{n=1}^{\infty}$ where $a_n = \frac{1}{2^n}$, we can make a_n arbitrarily close to 0 by taking n large enough. For example, if we want $\left|\frac{1}{2^n} - 0\right| < 0.01$, we can take $n > 7$; if we want $\left|\frac{1}{2^n} - 0\right| < 0.001$, we can take $n > 10$ etc.

Intuitively, Condition ($*$) means that if we let n increase without bound (or let n approach "∞", an imaginary point very far on the right), the value a_n approaches L. Geometrically, this means that the point (n, a_n) approaches the horizontal line $y = L$ as n increases without bound.

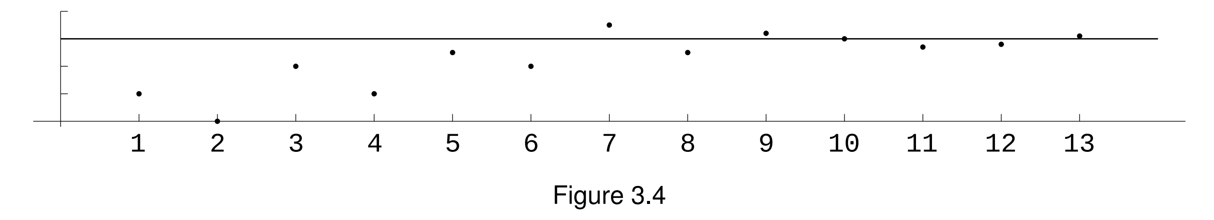


Figure 3.4

Remark For simplicity, instead of saying Condition ($*$), we will say

($**$) a_n is close to L if n is large.

In the definition of "*convergent*", it is clear that if L exists, then it is unique. We say that L is the *limit* of $(a_n)_{n=1}^{\infty}$ and we write $\lim_{n\to\infty} a_n = L$.

FAQ Can we just write $\lim a_n = L$, omitting $n \to \infty$?

Answer For sequences, this will not cause ambiguity. However, for functions, we will consider (in later sections) *limits at infinity* as well as *limits at a point a* (where $a \in \mathbb{R}$). The notations $\lim_{x\to\infty} f(x)$ and $\lim_{x\to a} f(x)$ have different meanings. □

The following rules can be proved by definition using an alternative method, called ϵ-N method, to describe condition ($*$). However, the ϵ-N definition is outside the scope of this course. Readers may convince themselves that the rules are true using intuition.

Rules for Limits of Sequences

(L1) $\lim_{n\to\infty} k = k$ (where k is a constant)

(L2) $\lim_{n\to\infty} \frac{1}{n^p} = 0$ (where p is a positive constant)

(L3) $\lim_{n\to\infty} \frac{1}{b^n} = 0$ (where b is a constant greater than 1)

(L4) $\lim_{n\to\infty} (a_n + b_n) = \lim_{n\to\infty} a_n + \lim_{n\to\infty} b_n$

(L5) $\lim_{n\to\infty} a_n b_n = \lim_{n\to\infty} a_n \cdot \lim_{n\to\infty} b_n$

(L6) $\lim_{n\to\infty} \frac{a_n}{b_n} = \frac{\lim_{n\to\infty} a_n}{\lim_{n\to\infty} b_n}$ provided that $\lim_{n\to\infty} b_n \neq 0$.

Remark

- The meaning of (L1) is that if $a_n = k$ for all n where k is a constant, then the sequence $(a_n)_{n=1}^{\infty}$ is convergent and its limit is k.
- The meaning of (L4) is that if both $(a_n)_{n=1}^{\infty}$ and $(b_n)_{n=1}^{\infty}$ are convergent and their limits are L and M respectively, then $(a_n + b_n)_{n=1}^{\infty}$ is also convergent and its limit is $L + M$.
- The following is a special case of (L5). It can be obtained by putting $a_n = k$ for all n and applying (L1).

  (L5s) $\lim_{n\to\infty} kb_n = k \lim_{n\to\infty} b_n$
- Using (L4) and (L5s), we get

  (L4′) $\lim_{n\to\infty} (a_n - b_n) = \lim_{n\to\infty} a_n - \lim_{n\to\infty} b_n$

  In fact, Rule (L4) is valid for sum and difference of finitely many sequences. This general result will be referred to as Rule (L4). Similarly, the result for product of finitely many sequences will be referred to as Rule (L5).

In Problem 1 in the last section, the sequence obtained can be represented by the formula $a_n = 4 + \frac{1}{2^n}$. Our intuition tells us that the limit of the sequence is 4. Below we use rules for limits to justify this result.

Example Find $\lim_{n\to\infty} \left(4 + \frac{1}{2^n}\right)$, if it exists.

Explanation The sequence under consideration is given by $a_n = 4 + \frac{1}{2^n}$. The question asks for the following

(1) Does the limit of $(a_n)_{n=1}^{\infty}$ exist or not (or equivalently, is the sequence convergent)?

(2) If the answer to (1) is affirmative, find the limit.

Solution

$$\begin{aligned}
\lim_{n\to\infty} \left(4 + \frac{1}{2^n}\right) &= \lim_{n\to\infty} 4 + \lim_{n\to\infty} \frac{1}{2^n} && \text{Rule (L4)} \\
&= 4 + 0 && \text{Rules (L1) and (L3)} \\
&= 4
\end{aligned}$$

□

Remark Below is the logic in the above calculation:

(1) In the first step, because the constant sequence $(4)_{n=1}^{\infty}$ and the sequence $\left(\frac{1}{2^n}\right)_{n=1}^{\infty}$ are convergent, we can apply Rule (L4).

(2) The limits of the two sequences are found by Rule (L1) and Rule (L3) respectively in the second step.

The sequence in the next example is the one obtained in Problem 2 in the last section.

Example Find $\lim\limits_{n\to\infty} \dfrac{2n^3 - 3n^2 + n}{6n^3}$, if it exists.

Solution

$$\begin{aligned}
\lim_{n\to\infty} \frac{2n^3 - 3n^2 + n}{6n^3} &= \lim_{n\to\infty} \left(\frac{1}{3} - \frac{1}{2n} + \frac{1}{6n^2}\right) && \text{Rewrite the expression} \\
&= \lim_{n\to\infty} \frac{1}{3} - \lim_{n\to\infty} \left(\frac{1}{2} \cdot \frac{1}{n}\right) + \lim_{n\to\infty} \left(\frac{1}{6} \cdot \frac{1}{n^2}\right) && \text{Rule (L4), rewrite 2nd and 3rd terms} \\
&= \frac{1}{3} - \frac{1}{2} \cdot \lim_{n\to\infty} \frac{1}{n} + \frac{1}{6} \cdot \lim_{n\to\infty} \frac{1}{n^2} && \text{Rules (L1) and (L5s)} \\
&= \frac{1}{3} - \frac{1}{2} \cdot 0 + \frac{1}{6} \cdot 0 && \text{Rule (L2)} \\
&= \frac{1}{3}
\end{aligned}$$

□

Example Find $\lim\limits_{n\to\infty} (1 + 2n)$, if it exists.

Solution Limit does not exist. This is because we can't find any real number L satisfying the condition that $2n + 1$ is close to L if n is large. □

Remark If we apply rules for limits, we get

$$\begin{aligned}
\lim_{n\to\infty} (2n + 1) &= \lim_{n\to\infty} 2n + \lim_{n\to\infty} 1 && \text{Rule (L4)} \\
&= 2 \lim_{n\to\infty} n + 1 && \text{Rules (L1) and (L5s)}
\end{aligned}$$

However, we can't proceed because $\lim\limits_{n\to\infty} n$ does not exist. From this, we see that the given limit does not exist.

FAQ Can we say that $\lim\limits_{n\to\infty} (1 + 2n)$ is ∞?

Answer Limit of a sequence is a *real number* satisfying Condition $(*)$ given in the definition on page 76. Because ∞ is not a real number, we should say that the limit does not exist.

In the next section, we will discuss the meaning of $\lim\limits_{x\to\infty} f(x) = \infty$ etc. □

Example Find $\lim\limits_{n\to\infty} \dfrac{n + 1}{2n + 1}$, if it exists.

Explanation We can't use Rule (L6) because limits of the numerator and the denominator do not exist. However, we can't conclude from this that the given limit does not exist. To find the limit, we use a trick: *divide the numerator and the denominator by n.*

Solution

$$\begin{aligned}
\lim_{n\to\infty}\frac{n+1}{2n+1} &= \lim_{n\to\infty}\frac{\frac{n+1}{n}}{\frac{2n+1}{n}} && \text{Divide numerator and denominator by } n\\
&= \frac{\lim_{n\to\infty}(1+\frac{1}{n})}{\lim_{n\to\infty}(2+\frac{1}{n})} && \text{Rule (L6), rewrite numerator and denominator}\\
&= \frac{\lim_{n\to\infty}1+\lim_{n\to\infty}\frac{1}{n}}{\lim_{n\to\infty}2+\lim_{n\to\infty}\frac{1}{n}} && \text{Rule (L4)}\\
&= \frac{1+0}{2+0} && \text{Rules (L1) and (L2)}\\
&= \frac{1}{2}
\end{aligned}$$

□

Remark We can apply the following shortcut (called the *Leading Terms Rule*). The method is to throw away the constant term 1 in the numerator and the denominator *(note that if n is very large, compared with n or* 2*n,* 1 *is very small).*

$$\begin{aligned}
\lim_{n\to\infty}\frac{n+1}{2n+1} &= \lim_{n\to\infty}\frac{n}{2n}\\
&= \lim_{n\to\infty}\frac{1}{2} = \frac{1}{2}.
\end{aligned}$$

The Leading Terms Rule for limits of functions at infinity will be discussed in more details in the next section (see page 83).

Exercise 3.2

1. For each of the following, find the limit if it exists.

   (a) $\lim_{n\to\infty}\frac{1}{\sqrt{n}}$ (b) $\lim_{n\to\infty}(7-\frac{5}{n^2})$

   (c) $\lim_{n\to\infty}\frac{3n^2-4000}{2n^2+10000}$ (d) $\lim_{n\to\infty}\frac{n^2-12345}{n+1}$

   (e) $\lim_{n\to\infty}\frac{5n^2+4}{2n^3+3}$ (f) $\lim_{n\to\infty}\left(\frac{1}{n}+(-1)^n\right)$

2. Suppose \$50,000 is deposited at a bank and the annual interests rate is 2%.

   (a) What amount (correct to the nearest cent) will the account have after one year if interests is
      (i) compounded quarterly;
      (ii) compounded monthly?

   (b) If interest is compounded n times a year, express the amount A_n after one year in terms of n.

   *(c) Does $\lim_{n\to\infty}A_n$ exist? What is the value?

*3. For each of the following sequences $(a_n)_{n=1}^{\infty}$, use computer to find the first 100 (or more) terms. Does $\lim_{n\to\infty}a_n$ exists? If yes, what is the value?

   (a) $a_n=(1+\frac{1}{n})^n$

   (b) $a_n=(1+\frac{2}{n})^n$

   (c) $a_n=n\sin\frac{1}{n}$ (angles are in radians)

*4. Suppose (a_n) is a sequence such that $0 < a_n$ for all n and $a_1 > a_2 > a_3 > \cdots$. Does $\lim_{n\to\infty} a_n$ exist? What can you tell about the limit?

3.3 Limits of Functions at Infinity

When we consider limit of a sequence $(a_n)_{n=1}^{\infty}$, we let n approach ∞ through the discrete points $n = 1, 2, 3, \ldots$. Recall that a sequence means a sequence of real numbers; it is a function from $\mathbb{Z}_+$ to $\mathbb{R}$. In many cases, we will consider functions f from a subset of $\mathbb{R}$ to $\mathbb{R}$ such that $f(x)$ is defined when x is large. For such functions, we can let x approach ∞ continuously (through large real numbers) and consider the behavior of $f(x)$.

Convention A *function* means a function whose domain is a subset of $\mathbb{R}$ and whose codomain is $\mathbb{R}$, unless otherwise stated.

Definition Let f be a function such that $f(x)$ is defined for sufficiently large x. Suppose L is a real number satisfying the following condition:

$(*)$ $f(x)$ is arbitrarily close to L if x is sufficiently large.

Then we say that L is the *limit* of f *at infinity* and write $\lim_{x\to\infty} f(x) = L$.

Remark

- The condition "$f(x)$ *is defined for sufficiently large* x" means that there is a real number r such that $f(x)$ is defined for all $x > r$.
- L is called *the* limit because it is unique (if it exists).
- For simplicity, instead of saying Condition $(*)$, we will say

  $(**)$ $f(x)$ is close to L if x is large.

Condition $(*)$ means that if we let x increase without bound, then the value $f(x)$ approaches L. To visualize this, imagine a small creature living on the curve $y = f(x)$. Suppose the small creature moves to the right indefinitely. It will get "closer and closer" to the horizontal line $y = L$.

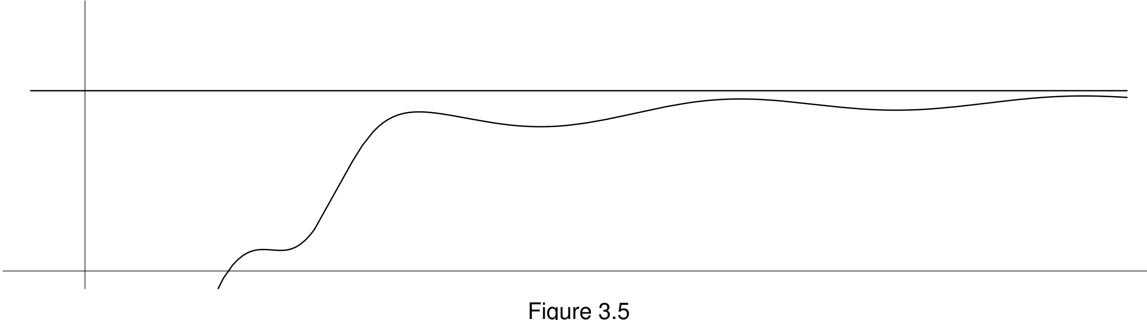


Figure 3.5

FAQ For sequences, we just say "*limit*". For functions, why are the words "*at infinity*" added?

Answer For functions, there are other types of limits. In Section 3.5, we will discuss limits of functions *at a* (where $a \in \mathbb{R}$). □

The following rules for limits of functions at infinity are similar to that for limits of sequences. In (4), (5), (5s) and (6), f and g are functions such that $f(x)$ and $g(x)$ are defined for sufficiently large x.

Rules for Limits of Functions at Infinity

(L1) $\lim_{x\to\infty} k = k$ (where k is a constant)

(L2) $\lim_{x\to\infty} \frac{1}{x^p} = 0$ (where p is a positive constant)

(L3) $\lim_{x\to\infty} \frac{1}{b^x} = 0$ (where b is a constant greater than 1)

(L4) $\lim_{x\to\infty} \big(f(x) \pm g(x)\big) = \lim_{x\to\infty} f(x) \pm \lim_{x\to\infty} g(x)$

The result is valid for sum and difference of finitely many functions.

(L5) $\lim_{x\to\infty} \big(f(x) \cdot g(x)\big) = \lim_{x\to\infty} f(x) \cdot \lim_{x\to\infty} g(x)$

The result is valid for product of finitely many functions.

(L5s) $\lim_{x\to\infty} \big(k \cdot g(x)\big) = k \cdot \lim_{x\to\infty} g(x)$

(L6) $\lim_{x\to\infty} \frac{f(x)}{g(x)} = \frac{\lim_{x\to\infty} f(x)}{\lim_{x\to\infty} g(x)}$ provided that $\lim_{x\to\infty} g(x) \neq 0$.

To consider limits of functions at infinity, we should first check the domains of the functions. For example, if $f(x) = \sqrt{1-x}$, the domain of f is $\{x \in \mathbb{R} : 1 - x \geq 0\} = (-\infty, 1]$; it is meaningless to talk about limit of f at infinity. In the next example, the domain of the function $1 - \frac{2}{x^3}$ is $\mathbb{R} \setminus \{0\}$; the function is defined for large x and hence we may consider its limit at infinity (whether the limit exists; and if exists, find the value).

Example Find $\lim_{x\to\infty} \left(1 - \frac{2}{x^3}\right)$, if it exists.

Solution

$$
\begin{aligned}
\lim_{x\to\infty} \left(1 - \frac{2}{x^3}\right) &= \lim_{x\to\infty} 1 - \lim_{x\to\infty} \left(2 \cdot \frac{1}{x^3}\right) && \text{Rule (L4), rewrite 2nd term} \\
&= 1 - 2 \cdot \lim_{x\to\infty} \frac{1}{x^3} && \text{Rules (L1) and (L5s)} \\
&= 1 - 2 \cdot 0 && \text{Rule (L2)} \\
&= 1
\end{aligned}
$$

□

Example Find $\lim_{x\to\infty} (2^{-x} + 3)$, if it exists.

Solution

$$
\begin{aligned}
\lim_{x\to\infty} (2^{-x} + 3) &= \lim_{x\to\infty} 2^{-x} + \lim_{x\to\infty} 3 && \text{Rule (L4)} \\
&= \lim_{x\to\infty} \frac{1}{2^x} + 3 && \text{Rewrite first term and Rule (L1)} \\
&= 0 + 3 && \text{Rule (L3)} \\
&= 3
\end{aligned}
$$

□

Example Find $\lim\limits_{x\to\infty} \dfrac{x^2+1}{3x^3-4x+5}$, if it exists.

Explanation Because limits of the numerator and denominator do not exist, we can't apply Rule (6). The first step is to divide the numerator and denominator by x^3 so that the limits at infinity of the new numerator and denominator exist.

Solution

$$\begin{aligned}
\lim_{x\to\infty} \frac{x^2+1}{3x^3-4x+5} &= \lim_{x\to\infty} \frac{\frac{x^2+1}{x^3}}{\frac{3x^3-4x+5}{x^3}} && \text{Divide numerator and denominator by } x^3 \\
&= \frac{\lim\limits_{x\to\infty}\left(\frac{1}{x}+\frac{1}{x^3}\right)}{\lim\limits_{x\to\infty}\left(3-\frac{4}{x^2}+\frac{5}{x^3}\right)} && \text{Rule (L6), rewrite numerator and denominator} \\
&= \frac{\lim\limits_{x\to\infty}\frac{1}{x}+\lim\limits_{x\to\infty}\frac{1}{x^3}}{\lim\limits_{x\to\infty}3-\lim\limits_{x\to\infty}\frac{4}{x^2}+\lim\limits_{x\to\infty}\frac{5}{x^3}} && \text{Rule (L4)} \\
&= \frac{0+0}{3-0+0} && \text{Rules (L1), (L2) and (L5s)} \\
&= 0
\end{aligned}$$

□

The next example is similar to the last one. To find limits at infinity for rational functions, we can divide the numerator and denominator by a suitable power of x.

Example Find $\lim\limits_{x\to\infty} \dfrac{x^3+1}{3x^3-4x+5}$, if it exists.

Solution

$$\begin{aligned}
\lim_{x\to\infty} \frac{x^3+1}{3x^3-4x+5} &= \lim_{x\to\infty} \frac{\frac{x^3+1}{x^3}}{\frac{3x^3-4x+5}{x^3}} && \text{Divide numerator and denominator by } x^3 \\
&= \frac{\lim\limits_{x\to\infty}\left(1+\frac{1}{x^3}\right)}{\lim\limits_{x\to\infty}\left(3-\frac{4}{x^2}+\frac{5}{x^3}\right)} && \text{Rule (L6), rewrite numerator and denominator} \\
&= \frac{\lim\limits_{x\to\infty}1+\lim\limits_{x\to\infty}\frac{1}{x^3}}{\lim\limits_{x\to\infty}3-\lim\limits_{x\to\infty}\frac{4}{x^2}+\frac{5}{x^3}} && \text{Rule (L4)} \\
&= \frac{1+0}{3-0+0} && \text{Rules (L1), (L2) and (L5s)} \\
&= \frac{1}{3}
\end{aligned}$$

□

To find limits at infinity for rational functions, we can also use the following shortcut.

Leading Terms Rule Let $f(x) = a_n x^n + a_{n-1}x^{n-1} + \cdots + a_1 x + a_0$ and $g(x) = b_m x^m + b_{m-1}x^{m-1} + \cdots + b_1 x + b_0$, where $a_n \neq 0$ and $b_m \neq 0$. Then we have

$$\lim_{x\to\infty} \frac{f(x)}{g(x)} = \lim_{x\to\infty} \frac{a_n x^n + a_{n-1}x^{n-1} + \cdots + a_1 x + a_0}{b_m x^m + b_{m-1}x^{m-1} + \cdots + b_1 x + b_0} = \lim_{x\to\infty} \frac{a_n x^n}{b_m x^m}$$

Proof The idea is to extract factor $a_n x^n$ in the numerator and $b_m x^m$ in the denominator. Putting

$$\varphi(x) = \frac{1 + \frac{a_{n-1}}{a_n}\cdot\frac{1}{x} + \frac{a_{n-2}}{a_n}\cdot\frac{1}{x^2} + \cdots + \frac{a_1}{a_n}\cdot\frac{1}{x^{n-1}} + \frac{a_0}{a_n}\cdot\frac{1}{x^n}}{1 + \frac{b_{m-1}}{b_m}\cdot\frac{1}{x} + \frac{b_{m-2}}{b_m}\cdot\frac{1}{x^2} + \cdots + \frac{b_1}{b_m}\cdot\frac{1}{x^{m-1}} + \frac{b_0}{b_m}\cdot\frac{1}{x^m}}$$

we have $\frac{f(x)}{g(x)} = \frac{a_n x^n}{b_m x^m}\cdot\varphi(x)$. It is straightforward to check that $\lim_{x\to\infty}\varphi(x) = 1$. Hence by Rule (5), we obtain the required result. □

Remark

- (a) If $n = m$, the limit is $\frac{a_n}{b_n}$.

  (b) If $n < m$, the limit is $\lim_{x\to\infty}\left(\frac{a_n}{b_m}\cdot\frac{1}{x^{m-n}}\right) = 0$.

  (c) If $n > m$, the limit is $\lim_{x\to\infty}\left(\frac{a_n}{b_m}\cdot x^{n-m}\right)$ which does not exist because as x increases indefinitely, x^{n-m} increases indefinitely.

- The Leading Terms Rule can also be applied to "functions similar to rational functions", for example, for $f(x) = x + 2\sqrt{x} + 3$ and $g(x) = 5x + 6\sqrt{x} + 7$, we have $\lim_{x\to\infty}\frac{x+2\sqrt{x}+3}{5x+6\sqrt{x}+7} = \lim_{x\to\infty}\frac{x}{5x}$

- The Leading Terms Rule can't be applied to limits of rational functions at a point: $\lim_{x\to a}\frac{f(x)}{g(x)}$, where $a \in \mathbb{R}$.

Below we re-do the last two examples using the Leading Terms Rule.

Example

$$\begin{aligned}
\lim_{x\to\infty}\frac{x^2+1}{3x^3-4x+5} &= \lim_{x\to\infty}\frac{x^2}{3x^3} && \text{Leading Terms Rule}\\
&= \lim_{x\to\infty}\left(\frac{1}{3}\cdot\frac{1}{x}\right) && \text{Simplify and rewrite expression}\\
&= \frac{1}{3}\cdot 0 && \text{Rules (L2) and (L5s)}\\
&= 0
\end{aligned}$$

Example

$$\begin{aligned}
\lim_{x\to\infty}\frac{x^3+1}{3x^3-4x+5} &= \lim_{x\to\infty}\frac{x^3}{3x^3} && \text{Leading Terms Rule}\\
&= \lim_{x\to\infty}\frac{1}{3} && \text{Simplify expression}\\
&= \frac{1}{3} && \text{Rule (L1)}
\end{aligned}$$

In the next example, the function can be considered as a product or a quotient of two functions. However, we can't apply Rule (5) or (6) because limit at infinity of one of the functions does not exist. To find the limit, we need the following result.

Sandwich Theorem *Let f, g and h be functions such that $f(x)$, $g(x)$ and $h(x)$ are defined for sufficiently large x. Suppose that $f(x) \leq g(x) \leq h(x)$ if x is sufficiently large and that both $\lim_{x\to\infty} f(x)$ and $\lim_{x\to\infty} h(x)$ exist and are equal (with common limit denoted by L). Then we have $\lim_{x\to\infty} g(x) = L$.*

Remark The condition "$f(x) \leq g(x) \leq h(x)$ *if x is sufficiently large*" means that there is a real number r such that the inequalities are true for all $x > r$.

Example Find $\lim_{x\to\infty} \frac{\sin x}{x}$, if it exists.

Explanation The given function can be written as a product of two functions: $\sin x$ and $\frac{1}{x}$. For the second function, its limit at infinity is 0. However, for the first function, its limit at infinity does not exist. Thus we can't apply Rule (5).

Solution Since $-1 \leq \sin x \leq 1$ for all real numbers x, it follows that

$$\frac{-1}{x} \leq \frac{\sin x}{x} \leq \frac{1}{x} \quad \text{for all } x > 0.$$

Note that $\lim_{x\to\infty} \frac{-1}{x} = \lim_{x\to\infty} \frac{1}{x} = 0$. Thus by the Sandwich Theorem, we have $\lim_{x\to\infty} \frac{\sin x}{x} = 0$. □

Example Find $\lim_{x\to\infty} (1 + \log x)$, if it exists.

Remark Since $\lim_{x\to\infty} \log x$ does not exist, we can't apply Rule (L4).

Solution $\lim_{x\to\infty} (1 + \log x)$ does not exist. This is because if x increases without bound, so does $1 + \log x$. □

Infinite Limits

In the last example, although limit does not exist, we know that if x increases indefinitely, so does $(1 + \log x)$. In the limit notation $\lim_{x\to\infty}$, the symbol $x \to \infty$ indicates that "*x increases indefinitely*", or "*x approaches ∞*". Using the same idea, we also write $1 + \log x \to \infty$ which indicates that the value increases indefinitely (as x increases indefinitely). Putting $y = 1 + \log x$, we write $y \to \infty$ as $x \to \infty$. Concerning the graph of $y = f(x)$ in the coordinate plane, $x \to \infty$ means that x goes to the right indefinitely, approaching the point ∞ (an imaginary point on the right) and $y \to \infty$ means that y goes up indefinitely, approaching the point ∞ (an imaginary point at the top).

Notation Let f be a function such that $f(x)$ is defined for sufficiently large x. Suppose that

(∗) $f(x)$ is arbitrarily large if x is sufficiently large.

Then we write $\lim_{x\to\infty} f(x) = \infty$.

Remark

- Because ∞ is not a real number, $\lim_{x\to\infty} f(x) = \infty$ does not mean the limit exists. In fact, it indicates that the limit does not exist and explains why it does not exist.

- Instead of $\lim_{x\to\infty} f(x) = \infty$, we also write $f(x) \to \infty$ as $x \to \infty$.
- For simplicity, instead of saying Condition $(*)$, we will say

  $(**)$ $\quad f(x)$ is large if x is large.
- Similar to $\lim_{x\to\infty} f(x) = \infty$, we also have $\lim_{x\to\infty} f(x) = -\infty$ which means that

  $(**)$ $\quad f(x)$ is large negative if x is large.



Example (a) $\lim_{x\to\infty}(1 + x^2) = \infty$ (limit does not exist) (b) $\lim_{x\to\infty}(1 - x^2) = -\infty$ (limit does not exist)

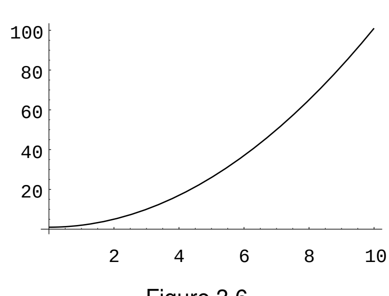


Figure 3.6

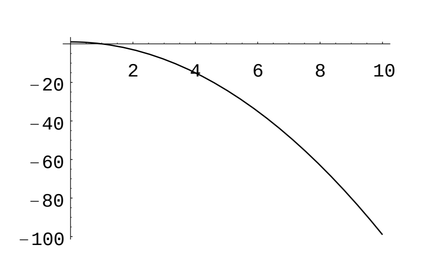
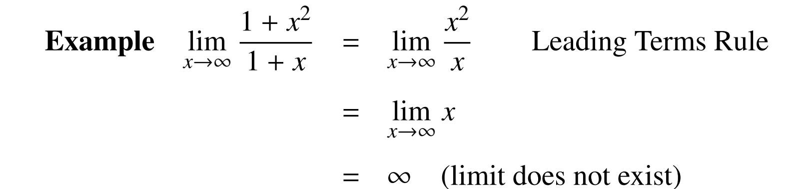


Figure 3.7



Example
$$\begin{aligned}\lim_{x\to\infty} \frac{1 + x^2}{1 + x} &= \lim_{x\to\infty} \frac{x^2}{x} \qquad \text{Leading Terms Rule}\\ &= \lim_{x\to\infty} x\\ &= \infty \quad \text{(limit does not exist)}\end{aligned}$$

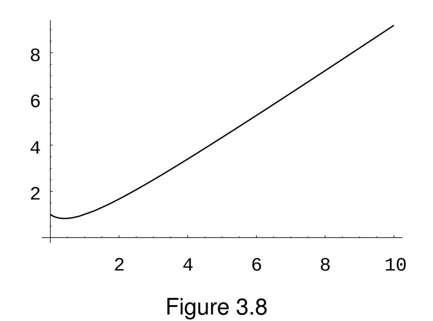


Figure 3.8

Limits at negative infinity

Similar to limits at infinity, we may consider *limits at negative infinity* provided that $f(x)$ is defined for x sufficiently large negative. Readers can figure out the meaning of the following notations:

- $\lim_{x\to-\infty} f(x) = L$ where L is a real number;
- $\lim_{x\to-\infty} f(x) = \infty$;
- $\lim_{x\to-\infty} f(x) = -\infty$.

Example

(1) $\lim_{x\to-\infty} \frac{1}{x} = 0$

(2) $\lim_{x\to-\infty} x^3 = -\infty$ (limit does not exist)

(3) $\lim_{x\to-\infty} (1 - x^3) = \infty$ (limit does not exist)

FAQ Can we perform addition, multiplication etc. with ∞ or $-\infty$?

Answer Yes and no. For example,

- $1 + \infty = \infty$,

- $2 \cdot \infty = \infty$.

However,

- $\infty - \infty$ is undefined,
- $0 \cdot \infty$ is undefined.

Be careful when you perform such operations. □

Exercise 3.3

1. For each of the following, find the limit if it exists.

   (a) $\displaystyle\lim_{x\to\infty} \frac{1}{\sqrt{x-1}}$ (b) $\displaystyle\lim_{x\to\infty} (15 - 16x^{-3})$

   (c) $\displaystyle\lim_{x\to\infty} 5^{-x}$ (d) $\displaystyle\lim_{x\to\infty} \sqrt{x}$

   (e) $\displaystyle\lim_{x\to\infty} \frac{x^2+9}{x^3+1}$ (f) $\displaystyle\lim_{x\to\infty} \frac{x^2+9}{x^2+1}$

   (g) $\displaystyle\lim_{x\to\infty} \frac{x^2+9}{x+1}$ (h) $\displaystyle\lim_{x\to\infty} \frac{|x|}{x}$

   (i) $\displaystyle\lim_{x\to-\infty} \frac{|x|}{x}$ (j) $\displaystyle\lim_{x\to-\infty} x\sin x$

2. The concentration C of a drug in a patient's bloodstream t hours after it was injected is given by

$$C(t) = \frac{0.15t}{t^2+3}.$$

   (a) Find $\displaystyle\lim_{t\to\infty} C(t)$.

   (b) Interpret the result in (a).

3. The population P of a certain small town t years from now is predicted to be $P(t) = 35000 + \dfrac{10000}{(t+2)^2}$.

   (a) Find the population in the long run.

   (b) Use computer to sketch the graph of P. What can you tell from the graph?

*4. For each of the following functions f, use computer to find $f(x)$ for large x. Guess whether $\displaystyle\lim_{x\to\infty} f(x)$ exists or not. If the limit exists, what is the limit?

   (a) $f(x) = \sqrt{x+1} - \sqrt{x}$

   (b) $f(x) = \sqrt{x^2+x} - x$

   (c) $f(x) = \dfrac{x^{99}}{2^x}$

3.4 One-sided Limits

In the last section, we consider limits at infinity (or negative infinity) by letting x approach the *imaginary point* ∞ (or $-\infty$). In this section, we consider limits at a point a on the real line by letting x approach a. Because x can approach a from the left-side or from the right-side, we have left-side and right-side limits. They are called *one-sided limits*.

Right-side Limits

Definition Let $a \in \mathbb{R}$ and let f be a function such that $f(x)$ is defined for x sufficiently close to and greater than a. Suppose L is a real number satisfying

(∗) $f(x)$ is arbitrarily close to L if x is sufficiently close to and greater than a.

Then we say that L is the *right-side limit* of f at a and we write $\lim\limits_{x \to a+} f(x) = L$.

Remark

- The condition "$f(x)$ *is defined for x sufficiently close to and greater than a*" means that there is a positive real number δ such that $f(x)$ is defined for all $x \in (a, a+\delta)$. For simplicity, instead of saying the condition, we will say "*f is defined on the right-side of a*".
- Instead of saying Condition (∗), we will say

  (∗∗) $f(x)$ is close to L if x is close to and greater than a.
- In the definition, it doesn't matter whether f is defined at a or not. If $f(a)$ is defined, its value has no effect on the existence and the value of $\lim\limits_{x \to a+} f(x)$. This is because right-side limit depends on the values of $f(x)$ for x close to and greater than a.

Example Let $f(x) = 1 - 2^{-\frac{1}{\sqrt{x}}}$. The domain of f is $(0, \infty)$ and so f is defined on the right-side of 0. The graph of f is shown in the following figure.

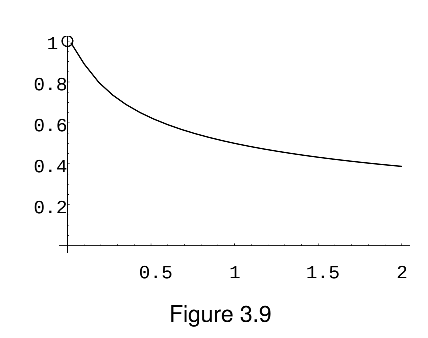


Figure 3.9

Imagine a small creature living on the curve. Suppose it moves to the left so that the x-coordinate of its position approaches 0 (from the right). From the graph, we see that the y-coordinate will approach 1. In other words, the right-side limit of f at 0 is 1, that is, $\lim\limits_{x \to 0+} (1 - 2^{-\frac{1}{\sqrt{x}}}) = 1$.

FAQ How can we know the graph of f ?

Answer This example is chosen to illustrate the idea of right-side limit. The graph is generated by computer.

- To see why $\lim\limits_{x \to 0+} (1-2^{-\frac{1}{\sqrt{x}}}) = 1$, note that if x is small positive, then so is $\sqrt{x}$ and hence $\frac{1}{\sqrt{x}}$ is large positive; the expression $2^{-\frac{1}{\sqrt{x}}} = 2^{-\text{large positive}} = \frac{1}{2^{\text{large positive}}}$ is small positive and so $f(x) = (1 - \text{small positive})$ is close to 1.
- To see why the graph goes down (as x increases), note that if x increases, then so does $\sqrt{x}$ and so $\frac{1}{\sqrt{x}}$ decreases; hence $-\frac{1}{\sqrt{x}}$ increases; therefore $2^{-\frac{1}{\sqrt{x}}}$ increases and thus $f(x)$ decreases. An alternative way to see this is to use differentiation (see Chapter 5 and Chapter 9). □

Example Let $f(x) = \sqrt{x}$. The domain of f is $[0, \infty)$ and so f is defined on the right-side of 0.

Note that if x is close to and greater than 0, then $f(x)$ is close to 0. This means that $\lim\limits_{x \to 0+} \sqrt{x} = 0$.

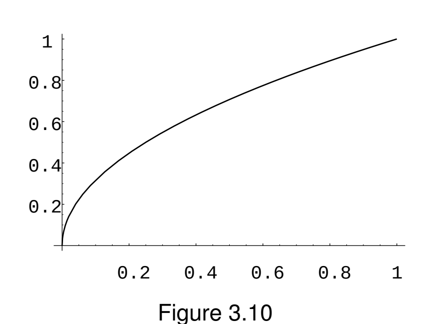


Figure 3.10

Example Let $f(x) = \sin\frac{1}{x}$. The domain of f is $\mathbb{R} \setminus \{0\}$. Thus f is defined on the right-side of 0 and we can consider the right-side limit of f at 0.

Remark In fact, $f(x)$ is also defined for $x < 0$ and so we can consider its left-side limit. See the definition for *left-side limit* below.

The graph of f (for $0 < x \leq 2$) is shown in Figure 3.11. Note that when x approaches 0 from the right-side, $f(x)$ oscillates between -1 and 1. Thus $\lim_{x\to 0+} \sin\frac{1}{x}$ does not exist.

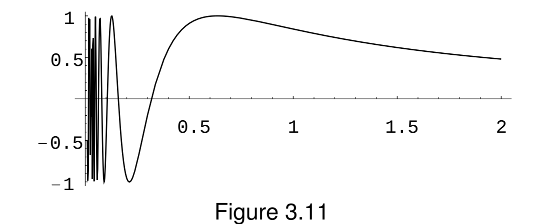


Figure 3.11

Example Let $f(x) = \frac{\sin x}{x^2 - x}$. The domain of f is $\mathbb{R} \setminus \{0, 1\}$. Thus $f(x)$ is defined for $0 < x < 1$ and we can consider $\lim_{x\to 0+} f(x)$.

The graph of f (for $0 < x \leq 0.8$) is shown in Figure 3.12. From the graph, we see that $\lim_{x\to 0+} \frac{\sin x}{x^2 - x} = -1$.

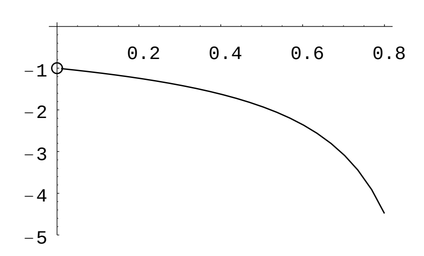


Figure 3.12

Remark The limit can be calculated using the following fact:

$$\lim_{x\to 0} \frac{\sin x}{x} = 1$$

It is a *two-sided limit*. See next section for more details.

Left-side Limits

If a function f is defined on the left-side of a, we can consider its left-side limit. The notation $\lim_{x\to a-} f(x) = L$ means that

(∗) $f(x)$ is arbitrarily close to L if x is sufficiently close to and less than a.

Example Let $f(x) = \frac{\sin x}{x^2 - x}$. The domain of f is $\mathbb{R} \setminus \{0, 1\}$. Thus $f(x)$ is defined for $x < 0$ and we can consider $\lim_{x\to 0-} f(x)$.

The graph of f (for x close to and less than 0) is shown in the following figure. From the graph, we see that $\lim_{x\to 0-} \frac{\sin x}{x^2 - x} = -1$.

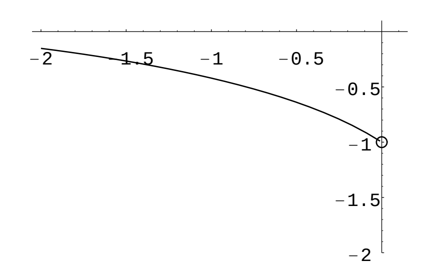


Figure 3.13

Similar to $\lim_{x\to\infty} f(x) = \infty$ etc., we have the following notations:

(1) $\lim_{x\to a+} f(x) = \infty$

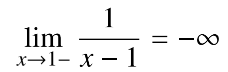
(2) $\lim_{x\to a+} f(x) = -\infty$

(3) $\lim_{x\to a-} f(x) = \infty$

(4) $\lim_{x\to a-} f(x) = -\infty$

Readers can figure out the meaning of the notations themselves. Geometrically, if any one of these notations is true, then the line $x = a$ is a *vertical asymptote* for the graph of f.

Example $\displaystyle\lim_{x\to 0+} \frac{1}{x} = \infty$



$\displaystyle\lim_{x\to 1-} \frac{1}{x-1} = -\infty$

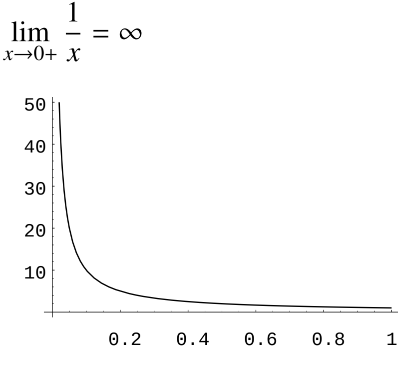


Figure 3.14

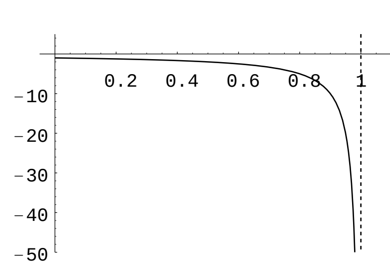


Figure 3.15

Exercise 3.4

1. For each of the following, find the limit if it exists.

   (a) $\displaystyle\lim_{x\to 1+} \sqrt{x-1}$ (b) $\displaystyle\lim_{x\to 0+} \frac{1}{\sqrt{x}}$

   (c) $\displaystyle\lim_{x\to 0-} \frac{1}{x^3}$ (d) $\displaystyle\lim_{x\to 0-} \sin\frac{1}{x}$

   (e) $\displaystyle\lim_{x\to 0+} (2 - 3^{\frac{1}{x}})$ (f) $\displaystyle\lim_{x\to 0-} (2 - 3^{\frac{1}{x}})$

3.5 Two-sided Limits

Definition Let $a \in \mathbb{R}$ and let f be a function that is defined on the left-side and right-side of a. Suppose that both $\displaystyle\lim_{x\to a-} f(x)$ and $\displaystyle\lim_{x\to a+} f(x)$ exist and are equal (with the common limit denoted by L which is a real number). Then the *two-sided limit*, or more simply, the *limit* of f at a is defined to be L, written $\displaystyle\lim_{x\to a} f(x) = L$.

Remark

- $\displaystyle\lim_{x\to a} f(x) = L$ means that

  $(*)$ $f(x)$ is arbitrarily close to L if x is sufficiently close to (but not equal to) a.

- The condition "*f be a function that is defined on the left-side and the right-side of a*" means that there is a positive real number δ such that $f(x)$ is defined for all $x \in (a-\delta) \cup (a, a+\delta)$.

- In considering $\displaystyle\lim_{x\to a} f(x)$, it doesn't matter whether f is defined at a or not.

Example Let $f(x) = \dfrac{\sin x}{x^2 - x}$. The domain of f is $\mathbb{R} \setminus \{0, 1\}$. Thus f is defined on the left-side and right-side of 0. In Section 3.4, the left-side and right-side limits of f at 0 were found to be

$$\lim_{x\to 0+} \frac{\sin x}{x^2 - x} = -1 \quad \text{and} \quad \lim_{x\to 0-} \frac{\sin x}{x^2 - x} = -1.$$

Thus, by definition, the limit of f at 0 exists and $\displaystyle\lim_{x\to 0} \frac{\sin x}{x^2 - x} = -1$.

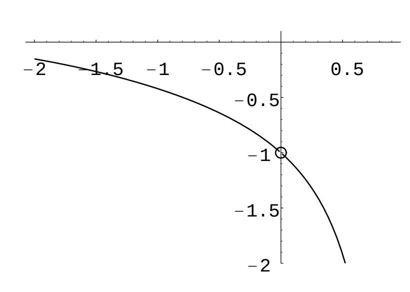


Figure 3.16

Example Let $f(x) = \dfrac{x}{|x|}$.

- For $x > 0$, we have $|x| = x$ and so $f(x) = \frac{x}{x} = 1$.
  Hence we get $\lim\limits_{x\to 0+} \frac{x}{|x|} = 1$.
- For $x < 0$, we have $|x| = -x$ and so $f(x) = \frac{x}{-x} = -1$.
  Hence we get $\lim\limits_{x\to 0-} \frac{x}{|x|} = -1$.

Therefore, $\lim\limits_{x\to 0} \frac{x}{|x|}$ does not exist.

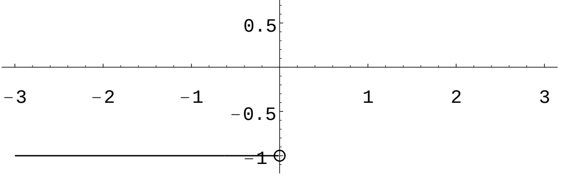


Figure 3.17

The following rules are useful to find limits of functions at a point a. In Rules (La4), (La5), (La5s) and (La6), f and g are functions that are defined on the left-side and right-side of a. Some of the rules are similar to that for limits of functions at infinity.

Rules for Limits of Functions at a Point

(La1) $\lim\limits_{x\to a} k = k$ (where $a \in \mathbb{R}$ and k is a constant)

(La2) $\lim\limits_{x\to a} x^n = a^n$ (where $a \in \mathbb{R}$ and n is a positive integer)

(La2′) $\lim\limits_{x\to a} \sqrt[n]{x} = \sqrt[n]{a}$ (where $a \in \mathbb{R}$ and n is an odd positive integer)

$\lim\limits_{x\to a} \sqrt[n]{x} = \sqrt[n]{a}$ (where $0 < a \in \mathbb{R}$ and n is an even positive integer)

(La3) $\lim\limits_{x\to a} b^x = b^a$ (where $a \in \mathbb{R}$ and b is a positive real number)

(La4) $\lim\limits_{x\to a} \big(f(x) \pm g(x)\big) = \lim\limits_{x\to a} f(x) \pm \lim\limits_{x\to a} g(x)$

The result is valid for sum and difference of finitely many functions.

(La5) $\lim\limits_{x\to a} \big(f(x) \cdot g(x)\big) = \lim\limits_{x\to a} f(x) \cdot \lim\limits_{x\to a} g(x)$

The result is valid for product of finitely many functions.

(La5s) $\lim\limits_{x\to a} \big(k \cdot g(x)\big) = k \cdot \lim\limits_{x\to a} g(x)$ (where k is a constant)

(La6) $\lim\limits_{x\to a} \dfrac{f(x)}{g(x)} = \dfrac{\lim\limits_{x\to a} f(x)}{\lim\limits_{x\to a} g(x)}$ provided that $\lim\limits_{x\to a} g(x) \neq 0$.

Example Find $\lim\limits_{x\to 4}(1 + x^2)$, if it exists.

Solution

$$\begin{aligned} \lim_{x\to 4}(1 + x^2) &= \lim_{x\to 4} 1 + \lim_{x\to 4} x^2 && \text{Rule (La4)} \\ &= 1 + 4^2 && \text{Rules (La1) and (La2)} \\ &= 17 \end{aligned}$$

□

Recall that a polynomial function p is a function that can be written in the following form

$$p(x) = c_n x^n + c_{n-1} x^{n-1} + \cdots + c_1 x + c_0,$$

where $c_0, c_1, \ldots, c_n$ are constants. Using the method in the above example, we can prove the following theorem which means that the limit of a polynomial function at any real number can be found by substitution.

Theorem 3.5.1 *Let $p(x)$ be a polynomial and let a be a real number. Then we have*

$$\lim_{x \to a} p(x) = p(a).$$

Example Find $\lim\limits_{x \to 2} \dfrac{x-1}{x^2+x-2}$, if it exists.

Solution

$$\begin{aligned} \lim_{x \to 2} \frac{x-1}{x^2+x-2} &= \frac{\lim\limits_{x \to 2}(x-1)}{\lim\limits_{x \to 2}(x^2+x-2)} && \text{Rule (La6)} \\ &= \frac{2-1}{2^2+2-2} && \text{Theorem 3.5.1} \\ &= \frac{1}{4} \end{aligned}$$

□

Recall that a rational function r is a function that can be written in the form

$$r(x) = \frac{p(x)}{q(x)},$$

where p and q are polynomial functions. Using the method in the above example, we can prove the following theorem which means that the limit of a rational function at any a belonging to its domain can be found by substitution.

Theorem 3.5.2 *Let $p(x)$ and $q(x)$ be polynomials and let a be a real number. Suppose that $q(a) \neq 0$. Then we have*

$$\lim_{x \to a} \frac{p(x)}{q(x)} = \frac{p(a)}{q(a)}.$$

Example Find $\lim\limits_{x \to 1} \dfrac{x-1}{x^2+x-2}$, if it exists.

Explanation The rational function $f(x) = \frac{x-1}{x^2+x-2}$ is undefined at $x = 1$. This means that 1 does not belong to the domain of f and so we can't apply Theorem 3.5.2. If we substitute $x = 1$ into the numerator and denominator, we get $\frac{0}{0}$. We say that the limit is in the *indeterminate form* $\frac{0}{0}$.

To find the limit, we replace f by a function g which coincides with f on the left-side and right-side of 1 such that the limit of g at 1 can be found by substitution (see the following figures).

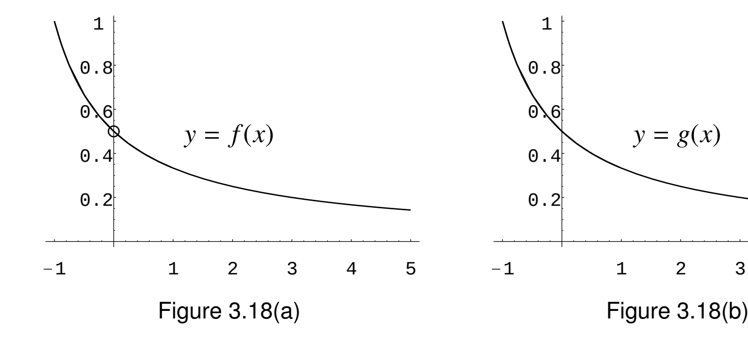


Figure 3.18(a) Figure 3.18(b)

The function g is given by $g(x) = \frac{1}{x+2}$. It can be found by simplifying the expression defining f:

$$f(x) = \frac{x-1}{(x-1)(x+2)} = g(x) \quad \text{for all } x \in \mathbb{R} \setminus \{1, -2\}.$$

Solution
$$\begin{aligned}
\lim_{x\to 1} \frac{x-1}{x^2+x-2} &= \lim_{x\to 1} \frac{x-1}{(x-1)(x+2)} \\
&= \lim_{x\to 1} \frac{1}{x+2} && \text{Simplify expression} \\
&= \frac{1}{1+2} && \text{Theorem 3.5.2} \\
&= \frac{1}{3}
\end{aligned}$$
□

Example Find $\lim_{x\to 1} \frac{x+1}{x^2+x-2}$, if it exists.

Explanation The rational function $f(x) = \frac{x+1}{x^2+x-2}$ is undefined at $x = 1$. Thus we can't use Theorem 3.5.2. If we put $x = 1$ into the numerator and denominator, we get $\frac{2}{0}$. Limits (left-side and right-side) of the form $\frac{\text{non-zero number}}{0}$ are ∞ or $-\infty$. See the solution below for more details.

Solution $\lim_{x\to 1} \frac{x+1}{x^2+x-2}$ does not exist. This is because if x is close to 1, the numerator is close to 2 whereas the denominator is close to 0 and so the fraction is very large in magnitude (may be positive or negative). □

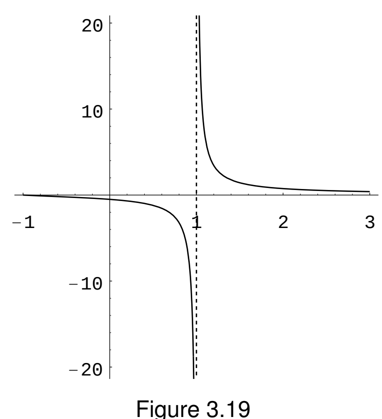


Figure 3.19

Remark Indeed, the denominator is $x^2 + x - 2 = (x-1)(x+2)$. Therefore, if x is close to and greater than 1, the denominator is small positive. Hence we have $\lim_{x\to 1+} \frac{x+1}{x^2+x-2} = \infty$. Similarly we have $\lim_{x\to 1-} \frac{x+1}{x^2+x-2} = -\infty$.

Example Let $f(x) = x^2 + 3$. Find $\lim_{h\to 0} \frac{f(x+h)-f(x)}{h}$.

Explanation The expression $\frac{f(x+h)-f(x)}{h}$ involves two variables x and h. However, the question asks for $\lim_{h\to 0}$ (limit of the expression as h approaches 0). This implies that x is considered as a constant. In this way, the expression $\frac{f(x+h)-f(x)}{h}$ is considered as a function of h, defined for all $h \neq 0$. The limit is in the indeterminate form $\frac{0}{0}$ because if we put $h = 0$ in the expression, the numerator and denominator are both 0. To find the limit, we simplify the expression so that the troublesome factor h in the denominator is canceled.

Solution
$$\begin{aligned}
\frac{f(x+h)-f(x)}{h} &= \frac{((x+h)^2+3)-(x^2+3)}{h} \\
&= \frac{(x^2+2xh+h^2+3)-(x^2+3)}{h} \\
&= \frac{2xh+h^2}{h} \\
&= \frac{h(2x+h)}{h} \\
&= 2x+h
\end{aligned}$$

Since $\dfrac{f(x+h)-f(x)}{h}$ and $2x+h$ (considered as functions of h) are equal on the left-side and right-side of $h=0$, it follows that

$$\begin{aligned}\lim_{h\to 0}\frac{f(x+h)-f(x)}{h} &= \lim_{h\to 0}(2x+h) \\ &= 2x+0 \qquad \text{Theorem 3.5.1} \\ &= 2x\end{aligned}$$

□

Remark The expression $\frac{f(x+h)-f(x)}{h}$ is called a *difference quotient*. Limits of difference quotients will be discussed in detail in Chapter 4.

Summary for Limits In this chapter, we have introduced the following types of limits:

$$\lim_{n\to\infty} a_n, \qquad \lim_{x\to\infty} f(x), \qquad \lim_{x\to-\infty} f(x), \qquad \lim_{x\to a-} f(x), \qquad \lim_{x\to a+} f(x), \qquad \lim_{x\to a} f(x).$$

Since the definitions for $\lim_{n\to\infty} a_n$ and $\lim_{x\to\infty} f(x)$ are similar, we will omit limits of sequences in the following discussion. Note that for functions, the five types of limits take the form

$$\lim_{x\to\square} f(x)$$

where □ can be ∞, $-\infty$, $a-$, $a+$ or a. The notation

$$\lim_{x\to\square} f(x) = L,$$

where L is a real number, means the following

$f(x)$ is arbitrarily close to L if x is "sufficiently close to (and different from)" □,

which, in short, is written as

$f(x)$ is close to L if x is "close to (and different from)" □,

where $f(x)$ *is close to* L has the usual meaning and

- x *is "close to (and different from)"* ∞ means x is large;
- x *is "close to (and different from)"* $-\infty$ means x is large negative;
- x *is "close to (and different from)"* $a-$ means x is close to and less than a;
- x *is "close to (and different from)"* $a+$ means x is close to and greater than a;
- x *is "close to (and different from)"* a has the usual meaning.

If we cannot find a real number L such that $\lim_{x\to\square} f(x) = L$, then we say that $\lim_{x\to\square} f(x)$ does not exist. There are several possibilities for this. We may have

$$\lim_{x\to\square} f(x) = \infty \quad \text{or} \quad \lim_{x\to\square} f(x) = -\infty \quad \text{or} \quad \text{other behavior such as oscillation.}$$

Similar to the above discussion, the notation $\lim_{x\to\square} f(x) = \infty$ means that $f(x)$ is "close to" ∞ if x is "close to (and different from)" □, where *"close to"* ∞ means large.

Exercise 3.5

1. For each of the following, find the limit, if it exists.

   (a) $\lim_{x\to 2}(x^2+3x-4)$ (b) $\lim_{x\to 7}(x^2-5x-8)^3$

   (c) $\lim_{x\to 0}\left(\frac{3x-5}{2x+7}\right)^2$ (d) $\lim_{x\to 1}\frac{x\sqrt{x^2+1}}{x+1}$

   (e) $\lim_{x\to -2} 2^{5+4x}$ (f) $\lim_{x\to -2}\frac{x^2-4}{x^2+x-2}$

   (g) $\lim_{x\to -2}\frac{x^2-4}{x^2-x+2}$ (h) $\lim_{x\to 6}\frac{x^2-6x}{x^2-5x-6}$

   (i) $\lim_{x\to 5}\frac{x^2+25}{x-5}$ (j) $\lim_{x\to 3}\frac{x-3}{x^2+9}$

   (k) $\lim_{x\to 4}\frac{\sqrt{x}-2}{x-4}$ (l) $\lim_{x\to 2}\frac{x-5}{2-\sqrt{x-1}}$

2. For each of the following f, find $\lim_{h\to 0}\frac{f(x+h)-f(x)}{h}$.

   (a) $f(x)=4x-13$ (b) $f(x)=x^3$

   (c) $f(x)=\frac{1}{x}$ (d) $f(x)=\sqrt{x}$

3.6 Continuous Functions

In the last section, we see that for "nice" functions, we can use substitution to find limits, that is, $\lim_{x\to a} f(x)$ equals $f(a)$. Functions with this property are called continuous functions. They are very important in the theory of more advanced calculus because they have many other nice properties.

Definition Let $a\in\mathbb{R}$ and let f be a function such that $f(x)$ is defined for x sufficiently close to a (including a). If the following condition holds,

$(*)\quad \lim_{x\to a} f(x)=f(a).$

then we say that f is *continuous* at a. Otherwise, we say that f is *discontinuous* (or *not continuous*) at a.

Remark

- The condition "$f(x)$ *is defined for x sufficiently close to a*" means that there exists a positive real number δ such that $f(x)$ is defined for all $x\in(a-\delta,a+\delta)$. This condition implies that f is defined on the left-side and right-side of a as well as at the point a. Condition $(*)$ means that the left-side and right-side limits exist and $\lim_{x\to a+} f(x)=\lim_{x\to a-} f(x)=f(a)$.
- Condition $(*)$ can be replaced by the following:

  $(*')\quad f(x)$ is arbitrarily close to $f(a)$ if x is sufficiently close to a.

  Many authors use $(*')$ as definition for "*f is continuous at a*".
- Since $\lim_{x\to a} x=a$, condition $(*)$ means that

  $(*'')\quad \lim_{x\to a} f(x)=f\left(\lim_{x\to a} x\right).$

  If we consider $f(x)$ as an operation: f acts on x, $(*'')$ means that the operation of taking f and that of taking limit commute, that is, the order of taking f and taking limit can be interchanged.

- Instead of saying $(*')$, we will say

  $(**)$ $f(x)$ is close to $f(a)$ if x is close to a.

  Roughly speaking, $(**)$ means that if x is change from a to $a+\Delta x$ where Δx is a small number (see the following figure which shows the graph of $y = f(x)$ where f is a function continuous at a), then the corresponding change in y, denoted by Δy, is small, where $\Delta y = f(a + \Delta x) - f(a)$.

  Remark Δx is a symbol to denote a small change in x; it doesn't mean a product of two numbers Δ and x.

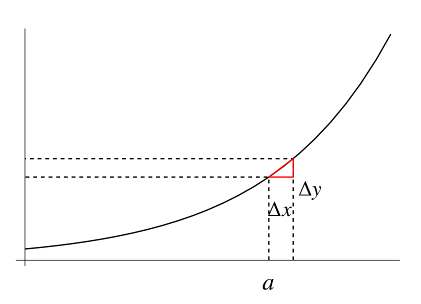


Figure 3.20

If a function f is undefined at a, it is meaningless to talk about whether f is continuous at a. Condition $(*)$ means that

(1) $\lim_{x\to a} f(x)$ exists;
(2) the limit in (1) equals $f(a)$.

If $\lim_{x\to a} f(x)$ does not exist or if $\lim_{x\to a} f(x)$ exists but does not equal $f(a)$, then f is discontinuous at a.

Example Let $f(x) = \begin{cases} -1 & \text{if } x < 0 \\ 0 & \text{if } x = 0 \\ 1 & \text{if } x > 0 \end{cases}$. Determine whether f is continuous at 0 or not.

Explanation The function is defined on the left-side and the right-side of 0 as well as at 0. Therefore, we may consider whether f is continuous at 0. In fact, we may consider whether f is continuous at 1 etc., but this is another question.

Solution By the definition of f, we have:

$$\lim_{x\to 0-} f(x) = \lim_{x\to 0-} -1 = -1 \quad \text{and} \quad \lim_{x\to 0+} f(x) = \lim_{x\to 0+} 1 = 1.$$

Since $\lim_{x\to 0-} f(x) \neq \lim_{x\to 0+} f(x)$, it follows that $\lim_{x\to 0} f(x)$ does not exist. Hence f is not continuous at 0.

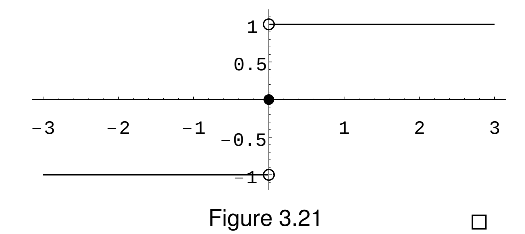


Figure 3.21 □

Example Let $f(x) = \begin{cases} x^2 & \text{if } x \neq 0, \\ 1 & \text{if } x = 0. \end{cases}$ For each real number a, determine whether f is continuous or discontinuous at a.

Explanation The domain of f is $\mathbb{R}$. So we may consider continuity of f at any point $a \in \mathbb{R}$ (that is, whether f is continuous at a).

Solution Consider the two cases where $a = 0$ or $a \neq 0$:

$(a = 0)$ Note that $f(x) = x^2$ on the left-side and the right-side of 0. Thus we have

$$\begin{aligned}
\lim_{x\to 0} f(x) &= \lim_{x\to 0} x^2 \\
&= 0^2 && \text{Theorem 3.5.1} \\
&= 0 \\
&\neq f(0)
\end{aligned}$$

Therefore, f is not continuous at 0.

$(a \neq 0)$ Note that $f(x) = x^2$ on the left-side and the right-side of a. Thus we have

$$\begin{aligned}
\lim_{x\to a} f(x) &= \lim_{x\to a} x^2 \\
&= a^2 && \text{Theorem 3.5.1} \\
&= f(a)
\end{aligned}$$

Therefore, f is continuous at a.

□

Remark The graph of f is shown in Figure 3.22.

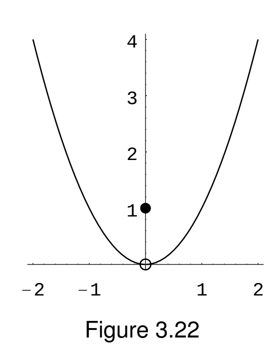


Figure 3.22

In the preceding definition, we consider continuity of a function f at a point a (a real number is considered as a point on the real line). In the next definition, we consider continuity of f on an open interval. Recall that an open interval is a subset of $\mathbb{R}$ that can be written in one of the following forms:

$$\begin{aligned}
(\alpha,\beta) &= \{x \in \mathbb{R} : \alpha < x < \beta\} \\
(\alpha,\infty) &= \{x \in \mathbb{R} : \alpha < x\} \\
(-\infty,\beta) &= \{x \in \mathbb{R} : x < \beta\} \\
(-\infty,\infty) &= \mathbb{R}
\end{aligned}$$

where α and β are real numbers, and for the first type, we need $\alpha < \beta$.

Definition Let I be an open interval and let f be a function defined on I. If f is continuous at every $a \in I$, then we say that f is *continuous on* I.

Remark

- In the definition, the condition "*f is a function defined on I*" means that f is a function such that $f(x)$ is defined for all $x \in I$, that is, $I \subseteq \operatorname{dom}(f)$.
- Since I is an open interval, we may consider continuity of f at any point a belonging to I.
- If there exists $a \in I$ such that f is not continuous at a, then f is *not continuous on I*.

Example In the last example, the domain of f is $\mathbb{R}$. The function is not continuous on $\mathbb{R}$ because it is not continuous at 0. For the open interval $(0, \infty)$, the function f is continuous at all a belonging to this interval. Therefore, f is continuous on $(0, \infty)$. Similarly, f is continuous on $(-\infty, 0)$.

Example Let $f(x) = \dfrac{1}{x}$. Show that f is continuous on $(0, \infty)$ as well as on $(-\infty, 0)$.

Explanation The domain of f is $\mathbb{R} \setminus \{0\}$. Since f is undefined at 0, we can't consider continuity of f at 0. The domain can be written as the union of two open intervals: $(-\infty, 0)$ and $(0, \infty)$. The question is to show that f is continuous on each of these two intervals, that is, f is continuous at every a in the two intervals. We may also say that f is continuous on $(-\infty, 0) \cup (0, \infty)$. However, this terminology will not be used in this course. We will consider continuity on intervals only, because functions continuous on (closed and bounded) intervals have nice properties (see Intermediate Value Theorem and Extreme Value Theorem below).

Proof For every $a \in (0, \infty)$, we have

$$
\begin{aligned}
\lim_{x\to a} f(x) &= \lim_{x\to a} \frac{1}{x} \\
&= \frac{1}{a} && \text{Theorem 3.5.2} \\
&= f(a)
\end{aligned}
$$

Therefore, f is continuous at a. By definition, f is continuous on $(0, \infty)$. Similarly, f is continuous on $(-\infty, 0)$. □

Remark The graph of f is shown in Figure 3.23.

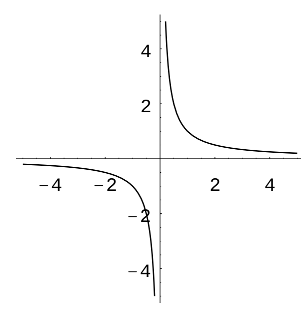


Figure 3.23

Remark Geometrically, a function f is continuous on an open interval I means that the graph of f on I has no "break"; if we use a pen to draw the graph on paper, we can draw it continuously without raising the pen above the paper.

The following two results give examples of continuous functions. They are just immediate consequences of the corresponding results for limits.

Theorem 3.6.1 *Every polynomial function is continuous on* $\mathbb{R}$.

Explanation The result means that if p is a polynomial function, then it is continuous on $\mathbb{R}$.

Proof Let p be a polynomial function. For every $a \in \mathbb{R}$, by Theorem 3.5.1, we have $\lim_{x\to a} p(x) = p(a)$, that is, p is continuous at a. Thus by definition, p is continuous on $\mathbb{R}$. □

Theorem 3.6.2 *Every rational function is continuous on every open interval contained in its domain.*

Explanation The result means that if f is a rational function and if I is an open interval with $I \subseteq \operatorname{dom}(f)$, then f is continuous on I. Recall that f can be written in the form $f(x) = \frac{p(x)}{q(x)}$ where $p(x)$ and $q(x)$ are polynomials.

- If $q(x)$ is never 0, then $\operatorname{dom}(f) = \mathbb{R}$.
- If $q(x) = 0$ has solutions, then $\operatorname{dom}(f)$ is the union of finitely many open intervals:

$$\operatorname{dom}(f) = \mathbb{R} \setminus \{z_1, z_2, \ldots, z_{k-1}, z_k\} = (-\infty, z_1) \cup (z_1, z_2) \cup \cdots \cup (z_{k-1}, z_k) \cup (z_k, \infty),$$

  where $z_1, \ldots, z_k$ are the (distinct) solutions arranged in increasing order.

Proof Let f be a rational function, that is, $f(x) = \frac{p(x)}{q(x)}$ where $p(x)$ and $q(x)$ are polynomials. Let I be an open interval with $I \subseteq \operatorname{dom}(f)$. For every $a \in I$, we have $a \in \operatorname{dom}(f)$ and so by the definition of domain, we have $q(a) \neq 0$. Therefore, by Theorem 3.5.2, we have $\lim_{x \to a} f(x) = \frac{p(a)}{q(a)} = f(a)$, that is, f is continuous at a. Thus by definition, f is continuous on I. □

In the preceding definition, we consider continuity on open intervals. If the domain of a function f is in the form $[a, b)$, we cannot talk about continuity of f at a because f is not defined on the left-side of a. Since f is defined on the right-side of a, we may consider $\lim_{x \to a+} f(x)$ and also whether the right-side limit equals $f(a)$.

Definition Let a be a real number and let f be a function defined on the right-side of a as well as at a. If $\lim_{x \to a+} f(x) = f(a)$, then we say that f is *right-continuous* at a.

Example Let $f(x) = \sqrt{x}$. The domain of f is $[0, \infty)$. Using a rule similar to Rule (La2′), we get

$$\begin{aligned} \lim_{x \to 0+} f(x) &= \lim_{x \to 0+} \sqrt{x} \\ &= 0 \\ &= f(0). \end{aligned}$$

Therefore, f is right-continuous at 0.

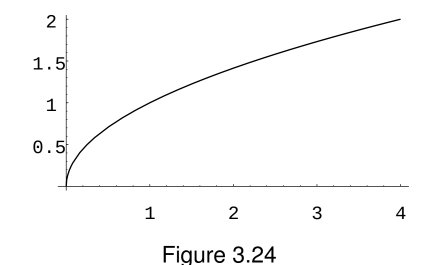


Figure 3.24

In the above example, f is also continuous at every $a > 0$. Thus, it is "continuous" at every a belonging to its domain, where "continuous at 0" means right-continuous at 0.

Definition Let I be an interval in the form $[c, d)$ where c is a real number and d is ∞ or a real number greater than c. Let f be a function defined on I. We say that f is *continuous on I* if it is continuous at every $a \in (c, d)$ and is right-continuous at c.

Similar to the above treatment, we may also consider continuity of functions f defined on intervals in the form $(c, d]$ or $[c, d]$.

Definition Let a be a real number and let f be a function defined on the left-side of a as well as at a. If $\lim_{x \to a-} f(x) = f(a)$, then we say that f is *left-continuous* at a.

Definition Let I be an interval in the form $(c, d]$ where d is a real number and c is $-\infty$ or a real number less than d. Let f be a function defined on I. We say that f is *continuous on I* if it is continuous at every $a \in (c, d)$ and is left-continuous at d.

Definition Let I be an interval in the form $[c, d]$ where c and d are real numbers and $c < d$. Let f be a function defined on I. We say that f is *continuous on I* if it is continuous at every $a \in (c, d)$ and is right-continuous at c and left-continuous at d.

Example Let $f : \mathbb{R} \longrightarrow \mathbb{R}$ be the function given by

$$f(x) = \begin{cases} |x| & \text{if } -1 \leq x \leq 1, \\ -1 & \text{otherwise.} \end{cases}$$

Discuss whether f is continuous on $[-1, 1]$.

Explanation In defining f, the word "*otherwise*" means that *if* $x < -1$ *or* $x > 1$; this is because it is given that $\operatorname{dom}(f) = \mathbb{R}$. Thus we have $f(x) = -1$ if $x < -1$ or $x > 1$.

Solution It is straightforward to check that f is continuous at every $a \in (-1, 1)$ and that f is left-continuous at 1 and right-continuous at -1. Thus by definition, f is continuous on $[-1, 1]$. □

Remark

- Note that f is also defined on the right-side of 1 (for example). Thus we can also consider the continuity of f at 1. In fact, since $\lim_{x \to 1+} f(x) = -1$ and $\lim_{x \to 1-} f(x) = 1$, it follows that $\lim_{x \to 1} f(x)$ does not exist and so f is not continuous at 1.
- Let I be an interval in the form $[c, d]$ or $[c, d)$ or $(c, d]$ and let f be a function defined on an *open interval containing I*. Then for every $a \in I$, we may consider whether f is continuous at a. The above example shows that f may be continuous on I but not continuous at some $a \in I$.

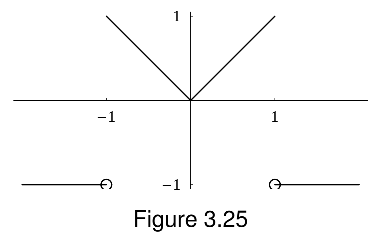


Figure 3.25

The following theorem describes an important property of continuous functions on intervals. The proof requires a deep understanding of *real numbers* and is beyond the scope of this course.

Intermediate Value Theorem *Let f be a function that is defined and continuous on an interval I. Then for every pair of elements a and b of I, and for every real number η between $f(a)$ and $f(b)$, there exists a number ξ between a and b such that $f(\xi) = \eta$.*

Explanation In the theorem, the condition "*f is a function that is defined and continuous on an interval I*" means that "*f is a function, I is an interval, $I \subseteq \operatorname{dom}(f)$ and f is continuous on I*".

- Let x, y and z be real numbers. We say that z *lies between* x and y if
  (1) $x \leq z \leq y$ for the case where $x \leq y$;
  (2) $y \leq z \leq x$ for the case where $y \leq x$.

  Note that if $x = y$, then z lies between x and y means that $z = x = y$.
- Because I is an interval, if a and b belong to I and $a < \xi < b$, then ξ belongs to I also.
- The result means that if f is a continuous function whose domain is an interval, then its range is either a singleton (in this case, f is a constant function) or an interval.

The following result is also called the *Intermediate Value Theorem*.

Corollary 3.6.3 *Let f be a function that is defined and continuous on an interval I. Suppose that a and b are elements of I such that $f(a)$ and $f(b)$ have opposite signs. Then there exists ξ between a and b such that $f(\xi) = 0$.*

Explanation The condition "*$f(a)$ and $f(b)$ have opposite signs*" means that one of the two values is positive and the other is negative.

Proof The result is a special case of the Intermediate Value Theorem. This is because $f(a)$ and $f(b)$ have opposite signs implies that 0 lies between $f(a)$ and $f(b)$. □

In the Intermediate Value Theorem, the assumption that f is continuous cannot be omitted. The following example is an illustration.

Example Let $f : [0, 2] \longrightarrow \mathbb{R}$ be defined by

$$f(x) = \begin{cases} -1 & \text{if } 0 \leq x \leq 1, \\ 1 & \text{if } 1 < x \leq 2. \end{cases}$$

Note that $f(0) = -1$ and $f(2) = 1$ have opposite signs. However, there does not exist any $\xi \in [0, 2]$ such that $f(\xi) = 0$.

We can't apply the Intermediate Value Theorem. This is because f is not continuous on $[0, 2]$. Indeed, it is not continuous at 0 since $\lim_{x\to 0-} f(x)$ and $\lim_{x\to 0+} f(x)$ are not equal.

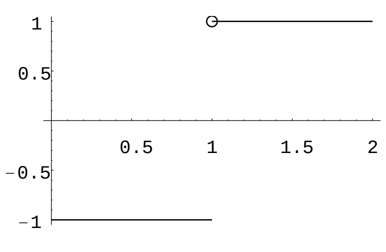


Figure 3.26

Corollary 3.6.4 *Let f be a function that is defined and continuous on an interval I. Suppose that f has no zero in I. Then f is either always positive in I or always negative in I.*

Explanation The condition "*f has no zero in I*" means that the equation $f(x) = 0$ has no solution in I, that is, $f(x) \neq 0$ for all $x \in I$. The conclusion "*f is either always positive in I or always negative in I*" means that either one of the following two cases is true:

(1) $f(x) > 0$ for all $x \in I$;

(2) $f(x) < 0$ for all $x \in I$.

Proof Suppose f takes both positive and negative values in I, that is, there exist $a, b \in I$ such that $f(a) < 0$ and $f(b) > 0$. Then by the Intermediate Value Theorem (Corollary 1), f has a zero between a and b which contradicts the assumption that f has no zero in I. □

The above corollary is also called the Intermediate Value Theorem. The following example illustrates how to apply the theorem to solve inequalities.

Example Find the solution set to the inequality $x^3 + 3x^2 - 4x - 12 \leq 0$.

Solution Let $p : \mathbb{R} \longrightarrow \mathbb{R}$ be the function given by

$$p(x) = x^3 + 3x^2 - 4x - 12.$$

Factorizing we get

$$p(x) = (x - 2)(x + 2)(x + 3).$$

The zeros of the function p are $-3, -2$ and 2 (and no more). Since p is continuous on $\mathbb{R}$, it follows from the Intermediate Value Theorem that on each of the following intervals, p is either always positive or always negative:

$$(-\infty, -3), \qquad (-3, -2), \qquad (-2, 2), \qquad (2, \infty).$$

To determine the sign of p on each of these intervals, we can just pick a point there and find the value (sign) of p at that point. Taking the points -4, -2.5, 0 and 3, we find that

$$p(-4) < 0, \qquad p(-2.5) > 0, \qquad p(0) < 0, \qquad p(3) > 0.$$

Thus we have

- $p(x) < 0$ for $x < -3$;
- $p(x) > 0$ for $-3 < x < -2$;
- $p(x) < 0$ for $-2 < x < 2$;
- $p(x) > 0$ for $x > 2$.

The solution set is $\{x \in \mathbb{R} : x \leq -3 \text{ or } -2 \leq x \leq 2\} = (-\infty, -3] \cup [-2, 2]$. □

Remark The above steps can be expressed in a compact form using a table:

	$x < -3$	$x = -3$	$-3 < x < -2$	$x = -2$	$-2 < x < 2$	$x = 2$	$x > 2$
$p(x)$	$-$ $p(-4) < 0$	0	$+$ $p(-2.5) > 0$	0	$-$ $p(0) < 0$	0	$+$ $p(3) > 0$

The next result describes an important property of functions continuous on closed and bounded intervals. It has many important consequences (for example, see the proof of the Mean Value Theorem in the appendix).

Extreme Value Theorem *Let f be a function that is defined and continuous on a closed and bounded interval $[a, b]$. Then f attains its maximum and minimum in $[a, b]$, that is, there exist $x_1, x_2 \in [a, b]$ such that*

$$f(x_1) \leq f(x) \leq f(x_2) \qquad \text{for all } x \in [a, b].$$

Explanation The theorem is a deep result. Its proof is beyond the scope of this course and is thus omitted.

The following two examples illustrate that in the Extreme Value Theorem,

- closed intervals cannot be replaced by open intervals;
- the assumption that f is continuous cannot be omitted.

Example Let $f : (0, 1) \longrightarrow \mathbb{R}$ be the function given by

$$f(x) = \frac{1}{x}$$

It is straightforward to show that f is continuous on $(0, 1)$. However, the function f does not attain its maximum nor minimum in $(0, 1)$. This is because the range of f is $(1, \infty)$; $f(x)$ can be arbitrarily large and it can be arbitrarily close to and greater than 1 but it can't be equal to 1.

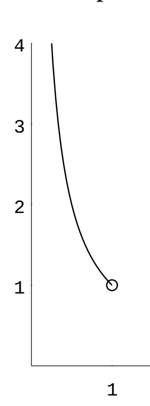


Figure 3.27

Example Let $f : [0, 1] \longrightarrow \mathbb{R}$ be the function given by

$$f(x) = \begin{cases} 1 & \text{if } x = 0, \\ \frac{1}{x} & \text{if } 0 < x \leq 1. \end{cases}$$

The function f does not attain its maximum in $[0, 1]$. This is because the range of f is $[1, \infty)$; $f(x)$ can be arbitrarily large.

Note that f is not right-continuous at 0 since $\lim_{x \to 0+} f(x) = \infty$ (limit does not exist).

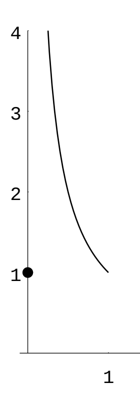


Figure 3.28

Exercise 3.6

1. Let $f(x) = \begin{cases} x^2 & \text{if } x < 1, \\ 1 & \text{if } 1 \leq x < 2, \\ \frac{1}{x} & \text{if } x \geq 2. \end{cases}$
   (a) Sketch the graph of f for $x \in [0, 5]$.
   (b) Find all the point(s) in $\mathbb{R}$ at which f is discontinuous.
2. Let $f(x) = \dfrac{x^2 + x - 2}{1 - \sqrt{x}}$.
   (a) What is the domain of f?
   (b) Find $\lim_{x \to 1} f(x)$.
   (c) Can we define $f(1)$ to make f continuous at 1? If yes, what is the value?
3. Let $f(x) = \sin \frac{1}{x}$.
   (a) What is the domain of f?
   (b) Find $\lim_{x \to 0} f(x)$.
   (c) Can we define $f(0)$ to make f continuous at 0? If yes, what is the value?
4. Let $p(x) = x^5 - x^4 - 5x^3 + x^2 + 8x + 4$. It is given that the equation $p(x) = 0$ has exactly two solutions, namely 2 and -1. Use this information to solve the inequality $p(x) > 0$.
5. Let $p(x) = x^5 - 6x^4 - 3x^3 + 5x^2 + 7$.
   (a) Show that the equation $p(x) = 0$ has a solution between 1 and 2.
   (b) It is given that $p(x) = 0$ has exactly one solution between 1 and 2. Is the solution closer to 1 or 2?

Chapter 4

Differentiation

4.1 Derivatives

Consider the curve shown Figure 4.1. It is clear from intuition that the "*slope*" changes as we move along the curve. At P', the slope is very steep whereas at P, the slope is gentle (in this sentence, slope means a piece of ground going up or down).

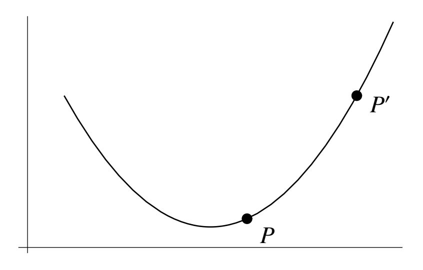


Figure 4.1

In elementary coordinate geometry, readers have learnt the concept "*slope of a line*". It is a number which measures how steep is the line. For a non-vertical line, its slope is given by

$$\frac{y_2 - y_1}{x_2 - x_1}$$

where (x_1, y_1) and (x_2, y_2) are two distinct points on the line and the value is independent of the choice of the two points.

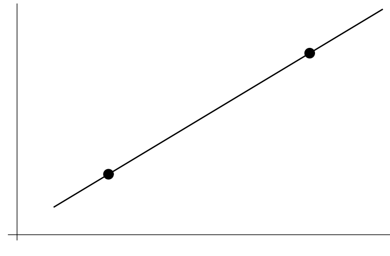


Figure 4.2

For curves, we shouldn't say "*slope of a curve*" because at different points of the curve, the slopes are different. Instead we should say "*slope of a curve at a point*". Below is how we define this concept.

First we have a curve $\mathcal{C}$ and a point P on the curve. To define the slope of $\mathcal{C}$ at P, take a point Q on the curve different from P. The line PQ is called a *secant line at P*. Its slope, denoted by m_{PQ}, can be found using the coordinates of P and Q. If we let Q move along the curve, the slope m_{PQ} changes.

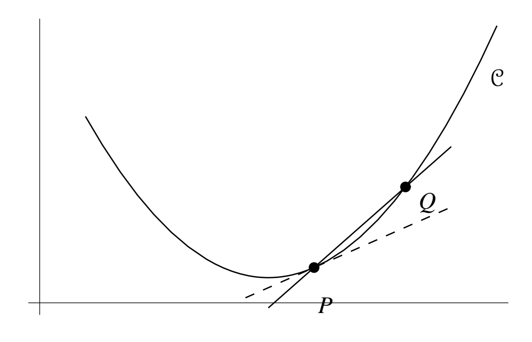


Figure 4.3

Suppose that as Q approaches P, the number m_{PQ} approaches a fixed value. This value, denoted by $m_{\mathcal{C},P}$ or simply m_P if the curve is understood, is called the *slope of* $\mathcal{C}$ *at P*; and the line with slope m_P and passing through P is called the *tangent line to the curve* $\mathcal{C}$ *at P*.

Remark The number $m_{\mathcal{C},P}$ (if exist) is the unique real number satisfying

(∗) m_{PQ} is arbitrarily close to $m_{\mathcal{C},P}$ if Q belonging to $\mathcal{C}$ is sufficiently close to (but different from) P.

In view of the concept "*limit of a function at a point*" and the notation $\lim_{x\to a} f(x)$, we may write

$$\lim_{\substack{Q\to P \\ \text{along } \mathcal{C}}} m_{PQ} = m_{\mathcal{C},P}$$

to mean that (∗) holds. Below, we will discuss how to find $\lim_{\substack{Q\to P \\ \text{along } \mathcal{C}}} m_{PQ}$ by rewriting it as the limit of a difference quotient.

Formula for Slope Suppose $\mathcal{C}$ is given by $y = f(x)$, where f is a function; and $P(x_0, f(x_0))$ is a point on $\mathcal{C}$. For any point Q on $\mathcal{C}$ with $Q \neq P$, its x-coordinate can be written as $x_0 + h$ where $h \neq 0$ (if $h > 0$, Q is on the right of P; if $h < 0$, Q is on the left of P). Thus, Q can be written as $(x_0 + h, f(x_0 + h))$. The slope m_{PQ} of the secant line PQ is

$$\begin{aligned} m_{PQ} &= \frac{f(x_0+h) - f(x_0)}{(x_0+h) - x_0} \\ &= \frac{f(x_0+h) - f(x_0)}{h} \end{aligned}$$

Note that as Q approaches P, the number h approaches 0. From these, we see that the slope of $\mathcal{C}$ at P (denoted by m_P) is

$$m_P = \lim_{h\to 0} \frac{f(x_0+h) - f(x_0)}{h} \tag{4.1.1}$$

provided that the limit exists.

Remark The limit in (4.1.1) is a two-sided limit. This is because Q can approach P from the left or from the right and so h can approach 0 from the left or from the right.

Example Find the slope of the curve given by $y = x^2$ at the point $P(3, 9)$.

Solution Put $f(x) = x^2$. By (4.1.1), the required slope (denoted by m_P) is

$$\begin{aligned} m_P &= \lim_{h\to 0} \frac{f(3+h) - f(3)}{h} \\ &= \lim_{h\to 0} \frac{(3+h)^2 - 3^2}{h} \\ &= \lim_{h\to 0} \frac{(9 + 6h + h^2) - 9}{h} \\ &= \lim_{h\to 0} \frac{6h + h^2}{h} \\ &= \lim_{h\to 0} (6 + h) \\ &= 6. \end{aligned}$$

Figure 4.4

□

Definition Let x_0 be a real number and let f be a function defined on an open interval containing x_0. Suppose the limit in (4.1.1) exists. Then we say that f is *differentiable* at x_0.

Convention Open intervals will be denoted by (a, b) including the cases where $a = -\infty$ and/or $b = \infty$, unless otherwise stated. Thus (a, b) can be any one of the following:

- (a, b) where $a, b \in \mathbb{R}$ and $a < b$;
- $(-\infty, b)$ where $a = -\infty$ and $b \in \mathbb{R}$;
- (a, ∞) where $a \in \mathbb{R}$ and $b = \infty$;
- $(-\infty, \infty)$ where $a = -\infty$ and $b = \infty$.

Remark

- The condition "*f is a function defined on an open interval containing x_0*" means that there is an open interval (a, b) such that $(a, b) \subseteq \operatorname{dom}(f)$ and $x_0 \in (a, b)$. Hence, f is defined on the left-side and right-side of x_0 as well as at x_0. The expression $\frac{f(x_0+h) - f(x_0)}{h}$ in the limit in (4.1.1), considered as a function of h, is defined on the left-side and the right-side of 0 but is undefined at 0.
- "*f is differentiable at x_0*" means that the slope of the curve $\mathcal{C}$ at P exists, where $\mathcal{C}$ is given by $y = f(x)$ and P is the point on $\mathcal{C}$ whose x-coordinate is x_0.
- There is an alternative way to describe the limit in (4.1.1). Putting $x = x_0 + h$, we have $x - x_0 = h$. Note that as h approaches to 0, x approaches to x_0. Hence we have

$$\lim_{h\to 0} \frac{f(x_0 + h) - f(x_0)}{h} = \lim_{x\to x_0} \frac{f(x) - f(x_0)}{x - x_0}.$$

Theorem 4.1.1 *Let x_0 be a real number and let f be a function defined on an open interval containing x_0. Suppose f is differentiable at x_0. Then f is continuous at x_0.*

Proof Since f is differentiable at x_0, by definition, $\lim_{x\to x_0} \frac{f(x)-f(x_0)}{x-x_0}$ exists (a real number). Hence we have

$$\begin{aligned}
\lim_{x\to x_0}\Big(f(x)-f(x_0)\Big) &= \lim_{x\to x_0}\left(\frac{f(x)-f(x_0)}{x-x_0}\cdot(x-x_0)\right) \\
&= \lim_{x\to x_0}\left(\frac{f(x)-f(x_0)}{x-x_0}\right)\cdot \lim_{x\to x_0}(x-x_0) && \text{Limit Rule (La5)} \\
&= \lim_{x\to x_0}\left(\frac{f(x)-f(x_0)}{x-x_0}\right)\cdot (x_0-x_0) && \text{Theorem 3.5.1} \\
&= \lim_{x\to x_0}\left(\frac{f(x)-f(x_0)}{x-x_0}\right)\cdot 0 \\
&= 0.
\end{aligned}$$

Therefore, we get

$$\begin{aligned}
\lim_{x\to x_0} f(x) &= \lim_{x\to x_0}\Big(f(x)-f(x_0)+f(x_0)\Big) \\
&= \lim_{x\to x_0}\Big(f(x)-f(x_0)\Big)+\lim_{x\to x_0} f(x_0) && \text{Limit Rule (La4)} \\
&= 0+f(x_0) && \text{From above and Limit Rule (La1)} \\
&= f(x_0).
\end{aligned}$$

That is, f is continuous at x_0. □

The following example illustrates that converse of Theorem 4.1.1 is not true.

Example Let $f(x) = |x|$. The domain of f is $\mathbb{R}$.

The function f is continuous at 0. This is because

- $\lim_{x\to 0-} f(x) = \lim_{x\to 0-}(-x) = 0$;
- $\lim_{x\to 0+} f(x) = \lim_{x\to 0+} x = 0$,

and so $\lim_{x\to 0} f(x) = 0 = f(0)$.

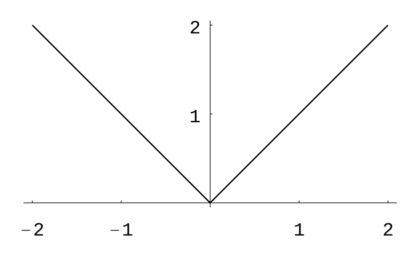


Figure 4.5

However, f is not differentiable at 0. This is because $\lim_{h\to 0}\frac{f(0+h)-f(0)}{h}$ does not exist as the left-side and right-side limits are unequal:

$$\begin{aligned}
\lim_{h\to 0+}\frac{f(0+h)-f(0)}{h} &= \lim_{h\to 0+}\frac{h-0}{h} \\
&= \lim_{h\to 0+} 1 \\
&= 1
\end{aligned}
\qquad\text{and}\qquad
\begin{aligned}
\lim_{h\to 0-}\frac{f(0+h)-f(0)}{h} &= \lim_{h\to 0-}\frac{-h-0}{h} \\
&= \lim_{h\to 0-} -1 \\
&= -1.
\end{aligned}$$

□

Definition Let f be a function.

(1) Suppose that f is differentiable at every point belonging to an open interval (a, b). Then we say that f is *differentiable on* (a, b).

(2) Suppose that f is differentiable at every point belonging to its domain. Then we say that f is a *differentiable function*.

Remark If f is a differentiable function, then its domain can be written as a union of open intervals.

Example Let $f(x) = |x|$. In a previous example, we have seen that f is not differentiable at 0. Thus f is not a differentiable function. Below we will show that f is differentiable at every $x_0 \neq 0$. Thus, f is differentiable on $(0, \infty)$ as well as on $(-\infty, 0)$.

$(x_0 > 0)$ In this case, if h is a small enough real number, then $x_0 + h > 0$ and so we have

$$\begin{aligned} f(x_0 + h) - f(x_0) &= |x_0 + h| - |x_0| \\ &= (x_0 + h) - x_0 = h, \end{aligned}$$

which yields

$$\begin{aligned} \lim_{h \to 0} \frac{f(x_0 + h) - f(x_0)}{h} &= \lim_{h \to 0} \frac{h}{h} \\ &= \lim_{h \to 0} 1 = 1. \end{aligned}$$

$(x_0 < 0)$ In this case, if h is a small enough real number, then $x_0 + h < 0$ and so we have

$$\begin{aligned} f(x_0 + h) - f(x_0) &= |x_0 + h| - |x_0| \\ &= -(x_0 + h) - (-x_0) = -h, \end{aligned}$$

which yields

$$\begin{aligned} \lim_{h \to 0} \frac{f(x_0 + h) - f(x_0)}{h} &= \lim_{h \to 0} \frac{-h}{h} \\ &= \lim_{h \to 0} -1 = -1. \end{aligned}$$

A function f is differentiable means that for every $x \in \operatorname{dom}(f)$, the limit of the difference quotient $\lim_{h \to 0} \frac{f(x + h) - f(x)}{h}$ exists; the limit is a real number and its value depends on x. In this way, we get a function, called the *derivative* of f and denoted by f', from $\operatorname{dom}(f)$ to $\mathbb{R}$.

The graph of f is a curve. The assumption that f is a differentiable function implies that at every point on the curve, the slope exists; at the point whose x-coordinate is x_0, the slope is $f'(x_0)$. Thus f' can be considered as *slope function*.

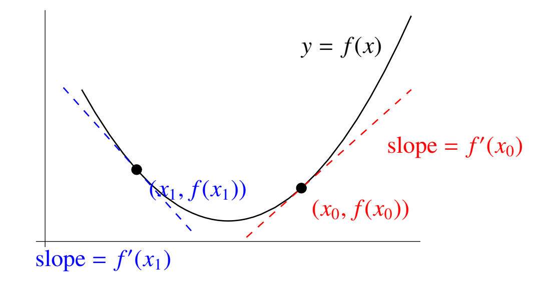


Figure 4.6

More generally, if f is differentiable only at some points in its domain, we can still define the derivative of f on a smaller set.

Definition Let f be a function that is differentiable at some points belonging in its domain. Then the *derivative* of f, denoted by f', is the function (from a subset of $\operatorname{dom}(f)$ into $\mathbb{R}$) given by

$$f'(x) = \lim_{h \to 0} \frac{f(x + h) - f(x)}{h},$$

where the domain of f' is $\left\{x \in \operatorname{dom}(f) : \lim_{h\to 0} \frac{f(x+h)-f(x)}{h} \text{ exists}\right\}$.

Example Let $f(x) = |x|$. Using results in a previous example, we see that

$$f'(x) = \begin{cases} 1 & \text{if } x > 0, \\ -1 & \text{if } x < 0, \end{cases}$$

where $\operatorname{dom}(f') = \{x \in \mathbb{R} : x \neq 0\}$.

Example Let $f(x) = x^2$. Find the derivative of f.

Explanation To find the derivative of f means to find the domain of f' and find a formula for $f'(x)$.

Solution By definition, we have

$$\begin{aligned} f'(x) &= \lim_{h\to 0} \frac{f(x+h)-f(x)}{h} \\ &= \lim_{h\to 0} \frac{(x+h)^2 - x^2}{h} \\ &= \lim_{h\to 0} \frac{(x^2+2xh+h^2) - x^2}{h} \\ &= \lim_{h\to 0} \frac{2xh+h^2}{h} \\ &= \lim_{h\to 0}(2x+h) \\ &= 2x. \end{aligned}$$

The domain of f' is $\mathbb{R}$. □

Remark The logic in the above solution is as follows:

(1) First we find that $\lim_{h\to 0} \frac{f(x+h)-f(x)}{h} = 2x$ for all $x \in \mathbb{R}$.

(2) From (1), we see that the domain of f' is $\mathbb{R}$ and $f'(x) = 2x$ for all $x \in \mathbb{R}$.

Example Let $f(x) = x^3$. Find $f'(x)$.

Solution By definition, we have

$$\begin{aligned} f'(x) &= \lim_{h\to 0} \frac{f(x+h)-f(x)}{h} \\ &= \lim_{h\to 0} \frac{(x+h)^3 - x^3}{h} \\ &= \lim_{h\to 0} \frac{(x^3+3x^2h+3xh^2+h^3) - x^3}{h} \\ &= \lim_{h\to 0} \frac{3x^2h+3xh^2+h^3}{h} \\ &= \lim_{h\to 0}(3x^2+3xh+h^2) \\ &= 3x^2. \end{aligned}$$

□

Remark The domain of f' is $\mathbb{R}$.

Terminology The process of finding derivatives is called *differentiation*.

In a previous example to find the slope of the parabola $y = f(x) = x^2$ at the point $(3, 9)$, we use definition to find $f'(3)$. In fact, if we know that $f'(x) = 2x$, then by direct substitution, we get $f'(3) = 6$. In the next section, we will discuss how to find $f'(x)$ using rules for differentiation.

Terminology $f'(x_0)$ is called the *derivative of f at x_0*.

To represent a function f, we sometimes write y or $f(x)$. Similarly, the derivative of a function can be represented in many ways.

Notation

- To denote the derivative of a function f, we have the following notations.

$$f', \quad y', \quad \frac{\mathrm{d}y}{\mathrm{d}x}, \quad Df, \quad Dy, \quad f'(x) \quad \text{and} \quad \frac{\mathrm{d}}{\mathrm{d}x}f(x).$$

- To denote the derivative of f at x_0, we have the following notations.

$$f'(x_0), \quad y'(x_0), \quad \left.\frac{\mathrm{d}y}{\mathrm{d}x}\right|_{x=x_0}, \quad Df(x_0) \quad \text{and} \quad Dy(x_0).$$

Some readers may wonder why we have the $'$ notation as well as the $\frac{\mathrm{d}}{\mathrm{d}x}$ notation. Calculus was "*invented*" by Newton and Leibniz independently in the late 17th century. Newton used $\dot{x}$ whereas Leibniz used $\frac{\mathrm{d}x}{\mathrm{d}t}$ to denote the derivative of x with respect to t (time). The notation y' is simple whereas $\frac{\mathrm{d}y}{\mathrm{d}x}$ reminds us that it is defined as a limit of *difference quotient*:

$$\frac{\mathrm{d}y}{\mathrm{d}x} = \lim_{\Delta x \to 0} \frac{\Delta y}{\Delta x}$$

where $\Delta x = h = (x + h) - x$ and $\Delta y = f(x + h) - f(x)$ are changes in x and y respectively.

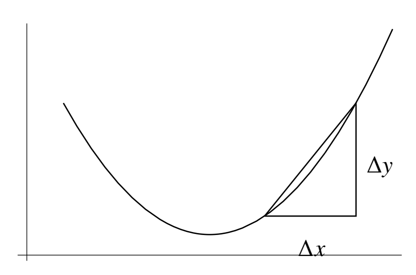


Figure 4.7

□

Caution $\frac{\mathrm{d}y}{\mathrm{d}x}$ is not a fraction.

- $\frac{\mathrm{d}y}{\mathrm{d}x}$ is f' or $f'(x)$; it is a function or an expression in x. It can be written as $\frac{\mathrm{d}}{\mathrm{d}x}y$ also. The notation $\frac{\mathrm{d}}{\mathrm{d}x}$ can be considered as an operation; to find $\frac{\mathrm{d}y}{\mathrm{d}x}$ means to perform the differentiation operation on y. Some authors use the notation Dy instead, where D stands for the differentiation operator (some authors use $D_x y$ to emphasize that the variable is x).
- Although we can define $\mathrm{d}y$ and $\mathrm{d}x$ (called *differentials*), $\frac{\mathrm{d}y}{\mathrm{d}x}$ does not mean "divide $\mathrm{d}y$ by $\mathrm{d}x$". In Chapter 10, we will describe *differentials* briefly (the purpose is to introduce the substitution method for integration).

Using $\frac{d}{dx}$ notation, the results obtained in the last two examples can be written as

$$\begin{aligned}\frac{d}{dx}x^2 &= 2x\\ \frac{d}{dx}x^3 &= 3x^2.\end{aligned}$$

In Exercise 3.5, Question 2(c) and (d), in terms of the $\frac{d}{dx}$ notation, the answers can be written as

$$\begin{aligned}\frac{d}{dx}x^{-1} &= -x^{-2}\\ \frac{d}{dx}x^{\frac{1}{2}} &= \frac{1}{2}x^{-\frac{1}{2}}.\end{aligned}$$

These are particular cases of a general result, called the *power rule* which will be discussed in Section 4.2.

Rate of Change

- The slope of a line is the rate of change of y with respect to x. The slope of a curve $y = f(x)$ at a point $P(x_0, f(x_0))$ is the limit of the slopes of the secant lines through P, so it is the rate of change of y with respect to x at P. That is, $f'(x_0)$ or $\left.\frac{dy}{dx}\right|_{x=x_0}$ is the rate of change of y with respect to x when $x = x_0$.

- If $x = t$ is time and $y = s(t)$ is the displacement function of a moving object then $s'(t_0)$ or $\left.\frac{ds}{dt}\right|_{t=t_0}$ is the rate of change of displacement with respect to time when $t = t_0$, that is, the (instantaneous) velocity at $t = t_0$. In the Introduction of Chapter 3, we consider the velocity of an object at time $t = 2$, where the displacement function is $s(t) = t^2$. Using differentiation, the velocity at $t = 2$ can be found easily:

$$\begin{aligned}s'(2) &= \left.\frac{d}{dt}t^2\right|_{t=2}\\ &= \left.2t\right|_{t=2}\\ &= 4.\end{aligned}$$

Remark The notation $\left.2t\right|_{t=2}$ means substitute $t = 2$ into the expression $2t$. More generally, the notation $\left.f(x)\right|_{x=x_0}$ means $f(x_0)$.

Exercise 4.1

1. For each of the following f, use definition to find $f'(x)$.

   (a) $f(x) = 2x^2 + 1$ (b) $f(x) = x^3 - 3x$

   (c) $f(x) = x^4$ (d) $f(x) = \frac{1}{x^2}$

4.2 Rules for Differentiation

Derivative of Constant *The derivative of a constant function is* 0 *(the zero function), that is*

$$\frac{d}{dx}c = 0,$$

where c is a constant.

Explanation In the above formula, we use c to denote the constant function f given by $f(x) = c$. The domain of f is $\mathbb{R}$. The result means that f is differentiable on $\mathbb{R}$ and that $f'(x) = 0$ for all $x \in \mathbb{R}$.

Proof Let $f : \mathbb{R} \longrightarrow \mathbb{R}$ be the function given by $f(x) = c$ for $x \in \mathbb{R}$. By definition, we get

$$\begin{aligned} f'(x) &= \lim_{h\to 0} \frac{f(x+h) - f(x)}{h} \\ &= \lim_{h\to 0} \frac{c - c}{h} \\ &= \lim_{h\to 0} \frac{0}{h} \\ &= \lim_{h\to 0} 0 \\ &= 0 \end{aligned}$$

□

Geometric meaning The graph of the constant function c is the horizontal line given by $y = c$. At every point on the line, the slope is 0.

Derivative of Identity Function *The derivative of the identity function is the constant function* 1, *that is*

$$\frac{\mathrm{d}}{\mathrm{d}x} x = 1.$$

Explanation In the above formula, we use x to denote the *identity function*, that is, the function f given by $f(x) = x$. The domain of f is $\mathbb{R}$. The result means that f is differentiable on $\mathbb{R}$ and $f'(x) = 1$ for all $x \in \mathbb{R}$.

Proof Let $f : \mathbb{R} \longrightarrow \mathbb{R}$ be the function given by $f(x) = x$ for $x \in \mathbb{R}$. By definition, we get

$$\begin{aligned} f'(x) &= \lim_{h\to 0} \frac{f(x+h) - f(x)}{h} \\ &= \lim_{h\to 0} \frac{x + h - x}{h} \\ &= \lim_{h\to 0} \frac{h}{h} \\ &= \lim_{h\to 0} 1 \\ &= 1 \end{aligned}$$

□

Geometric meaning The graph of the identity function is the line given by $y = x$. At every point on the line, the slope is 1.

Power Rule for Differentiation (positive integer version) *Let n be a positive integer. Then the power function x^n is differentiable on* $\mathbb{R}$ *and we have*

$$\frac{\mathrm{d}}{\mathrm{d}x} x^n = nx^{n-1}.$$

Explanation In the above formula, we use x^n to denote the n-th *power function*, that is, the function f given by $f(x) = x^n$. The domain of f is $\mathbb{R}$. The result means that f is differentiable on $\mathbb{R}$ and $f'(x) = nx^{n-1}$ for all

$x \in \mathbb{R}$. When $n = 1$, the formula becomes $\frac{\mathrm{d}}{\mathrm{d}x}x = 1x^0$. In the expression on the right side, x^0 is understood to be the constant function 1 and so the formula reduces to $\frac{\mathrm{d}}{\mathrm{d}x}x = 1$ which is the rule for derivative of the identity function. To prove that the result is true for all positive integers n, we can use mathematical induction. For base step, we know that the result is true when $n = 1$. For the induction step, we can apply product rule (will be discussed later). Below we will give alternative proofs for the power rule (positive integer version).

Proof Let $f : \mathbb{R} \longrightarrow \mathbb{R}$ be the function given by $f(x) = x^n$ for $x \in \mathbb{R}$. By definition, we get

$$\begin{aligned} f'(x) &= \lim_{h\to 0}\frac{f(x+h)-f(x)}{h} \\ &= \lim_{h\to 0}\frac{(x+h)^n - x^n}{h}. \end{aligned}$$

To find the limit, we "simplify" the numerator to obtain a factor h and then cancel it with the factor h in the denominator. For this, we can use:

- a *factorization formula* $a^n - b^n = (a-b)(a^{n-1} + a^{n-2}b + \cdots + ab^{n-2} + b^{n-1})$
- the *Binomial Theorem* $(a+b)^n = a^n + \binom{n}{1}a^{n-1}b + \binom{n}{2}a^{n-2}b^2 + \cdots + \binom{n}{n-2}a^2b^{n-2} + \binom{n}{n-1}ab^{n-1} + b^n$

  where $\binom{n}{1} = n$ and $\binom{n}{2} = \frac{n(n-1)}{2}$ and $\binom{n}{3} = \frac{n(n-1)(n-2)}{3\cdot 2}$ etc. are the binomial coefficients (see Binomial Theorem in the appendix). However, in the proof below, we only need to know the coefficient $\binom{n}{1} = n$; the other coefficients are not important.

(*Method 1*) By the factorization formula, we get

$$\begin{aligned} (x+h)^n - x^n &= (x+h-x)\Big((x+h)^{n-1} + (x+h)^{n-2}x + \cdots + (x+h)x^{n-2} + x^{n-1}\Big) \\ &= h\Big((x+h)^{n-1} + (x+h)^{n-2}x + \cdots + (x+h)x^{n-2} + x^{n-1}\Big) \end{aligned}$$

$$\begin{aligned} \text{Therefore,}\quad f'(x) &= \lim_{h\to 0}\Big((x+h)^{n-1} + (x+h)^{n-2}x + \cdots + (x+h)x^{n-2} + x^{n-1}\Big) \\ &= (x+0)^{n-1} + (x+0)^{n-2}\cdot x + \cdots + (x+0)\cdot x^{n-2} + x^{n-1} && \text{Theorem 3.5.1} \\ &= \underbrace{x^{n-1} + x^{n-1} + \cdots + x^{n-1} + x^{n-1}}_{n\text{ terms}} \\ &= nx^{n-1} \end{aligned}$$

(*Method 2*) From the Binomial Theorem, we get

$$\begin{aligned} (x+h)^n - x^n &= x^n + \binom{n}{1}x^{n-1}h + \binom{n}{2}x^{n-2}h^2 + \cdots + \binom{n}{n-2}x^2h^{n-2} + \binom{n}{n-1}xh^{n-1} + h^n - x^n \\ &= \binom{n}{1}x^{n-1}h + \binom{n}{2}x^{n-2}h^2 + \cdots + \binom{n}{n-2}x^2h^{n-2} + \binom{n}{n-1}xh^{n-1} + h^n \\ &= h\left(nx^{n-1} + \binom{n}{2}x^{n-2}h + \cdots + \binom{n}{n-2}x^2h^{n-3} + \binom{n}{n-1}xh^{n-2} + h^{n-1}\right) && \text{because } \binom{n}{1} = n \end{aligned}$$

$$\begin{aligned} \text{Therefore,}\quad f'(x) &= \lim_{h\to 0}\left(nx^{n-1} + \binom{n}{2}x^{n-2}h + \cdots + \binom{n}{n-2}x^2h^{n-3} + \binom{n}{n-1}xh^{n-2} + h^{n-1}\right) \\ &= nx^{n-1} + \binom{n}{2}x^{n-2}\cdot 0 + \cdots + \binom{n}{n-2}x^2\cdot 0 + \binom{n}{n-1}x\cdot 0 + 0 && \text{Theorem 3.5.1} \\ &= nx^{n-1}. \end{aligned}$$

□

Example Let $y = x^{123}$. Find $\dfrac{dy}{dx}$.

Explanation The notation $y = x^{123}$ represents a power function. To find $\frac{dy}{dx}$ means to find the derivative of the function.

Solution

$$\begin{aligned} \frac{dy}{dx} &= \frac{d}{dx}x^{123} \\ &= 123x^{123-1} && \text{Power Rule} \\ &= 123x^{122} \end{aligned}$$

□

Caution In the above solution, the first step is "*substitution*". It can also be written as $\frac{dx^{123}}{dx}$. It is wrong to write $\frac{dy}{dx}x^{123}$ which means $\frac{dy}{dx}$ multiplied by x^{123}. The notation $\frac{d}{dx}x^{123}$ is the derivative of the power function x^{123}. If we consider $\frac{d}{dx}$ as an operator, $\frac{d}{dx}x^{123}$ means "*perform the differentiation operation on the function x^{123}*".

Remark

- In the power rule, if we put $n = 0$, the left side is $\frac{d}{dx}x^0$ and right side is $0x^{-1}$. The function x^0 is understood to be the constant function 1. Thus $\frac{d}{dx}x^0 = 0$ by the rule for derivative of constant. The domain of the function x^{-1} is $\mathbb{R} \setminus \{0\}$. Therefore, $0x^{-1}$ is the function whose domain is $\mathbb{R} \setminus \{0\}$ and is always equal to 0. If we extend the domain to $\mathbb{R}$ and treat $0x^{-1}$ as the constant function 0, then the power rule is true when $n = 0$.
- Later in this section, we will show that the power rule is also true for negative integers n using the *quotient rule* (in fact, it is true for all real numbers n; the result will be called the *general power rule*).
- In Chapter 6, we will discuss integration which is the reverse process of differentiation. There is a result called the *power rule for integration*. However, in this chapter, "*power rule*" always means "*power rule for differentiation*".

Constant Multiple Rule for Differentiation *Let f be a function and let k be a constant. Suppose that f is differentiable. Then the function kf is also differentiable. Moreover, we have*

$$\frac{d}{dx}(kf)(x) = k \cdot \frac{d}{dx}f(x).$$

Explanation The function kf is defined by $(kf)(x) = k \cdot f(x)$ for $x \in \operatorname{dom}(f)$. The result means that if $f'(x)$ exists for all $x \in \operatorname{dom}(f)$, then $(kf)'(x) = k \cdot f'(x)$ for all $x \in \operatorname{dom}(f)$, that is, $(kf)' = k \cdot f'$.

Proof By definition, we have

$$\begin{aligned} (kf)'(x) &= \lim_{h\to 0} \frac{(kf)(x+h) - (kf)(x)}{h} \\ &= \lim_{h\to 0} \frac{k \cdot f(x+h) - k \cdot f(x)}{h} && \text{Definition of } kf \\ &= \lim_{h\to 0} \left(k \cdot \frac{f(x+h) - f(x)}{h} \right) \\ &= k \cdot \lim_{h\to 0} \frac{f(x+h) - f(x)}{h} && \text{Limit Rule (La5s)} \\ &= k \cdot f'(x) \end{aligned}$$

□

Remark There is a *pointwise version* of the *constant multiple rule*: the result remains valid if *differentiable function* is replaced by *differentiable at a point*.

Let f be a function and let k be a constant. Suppose that f is differentiable at x_0. Then the function kf is also differentiable at x_0. Moreover, we have $(kf)'(x_0) = k \cdot f'(x_0)$

There are also pointwise versions for the *sum rule*, *product rule* and *quotient rule* which will be discussed later. Readers can formulate the results themselves.

Example Let $y = 3x^4$. Find $\dfrac{\mathrm{d}y}{\mathrm{d}x}$.

Solution

$$\begin{aligned} \frac{\mathrm{d}y}{\mathrm{d}x} &= \frac{\mathrm{d}}{\mathrm{d}x}3x^4 \\ &= 3 \cdot \frac{\mathrm{d}}{\mathrm{d}x}x^4 && \text{Constant Multiple Rule} \\ &= 3 \cdot (4x^{4-1}) && \text{Power Rule} \\ &= 12x^3 \end{aligned}$$

□

Sum Rule for Differentiation (Term by Term Differentiation) *Let f and g be functions with the same domain. Suppose that f and g are differentiable. Then the function $f + g$ is also differentiable. Moreover, we have*

$$\frac{\mathrm{d}}{\mathrm{d}x}(f+g)(x) = \frac{\mathrm{d}}{\mathrm{d}x}f(x) + \frac{\mathrm{d}}{\mathrm{d}x}g(x).$$

Explanation The function $f + g$ is defined by $(f + g)(x) = f(x) + g(x)$ for x belonging to the common domain A of f and g. The result means that if $f'(x)$ and $g'(x)$ exist for all $x \in A$, then $(f + g)'(x) = f'(x) + g'(x)$ for all $x \in A$, that is, $(f + g)' = f' + g'$.

Proof By definition, we have

$$\begin{aligned} (f+g)'(x) &= \lim_{h\to 0} \frac{(f+g)(x+h) - (f+g)(x)}{h} \\ &= \lim_{h\to 0} \frac{\big(f(x+h) + g(x+h)\big) - \big(f(x) + g(x)\big)}{h} && \text{Definition of } f+g \\ &= \lim_{h\to 0} \frac{f(x+h) - f(x) + g(x+h) - g(x)}{h} \\ &= \lim_{h\to 0} \left(\frac{f(x+h) - f(x)}{h} + \frac{g(x+h) - g(x)}{h} \right) \\ &= \lim_{h\to 0} \frac{f(x+h) - f(x)}{h} + \lim_{h\to 0} \frac{g(x+h) - g(x)}{h} && \text{Limit Rule (La4)} \\ &= f'(x) + g'(x) \end{aligned}$$

□

Remark

- If the domains of f and g are not the same but their intersection is nonempty, we define $f + g$ to be the function with domain $\operatorname{dom}(f) \cap \operatorname{dom}(g)$ given by $(f + g)(x) = f(x) + g(x)$. The following is a more general version of the *sum rule*:

Let f and g be functions. Suppose that f and g are differentiable on an open interval (a, b). Then the function $f + g$ is also differentiable on (a, b). Moreover, on the interval (a, b), we have

$$\frac{d}{dx}(f+g)(x) = \frac{d}{dx}f(x) + \frac{d}{dx}g(x).$$

There are also more general versions for the *product rule* and *quotient rule*. Readers can formulate the results themselves.

- The result is also true for difference of two functions, that is,

$$\frac{d}{dx}(f-g)(x) = \frac{d}{dx}f(x) - \frac{d}{dx}g(x).$$

  The result for difference can be proved similar to that for sum. Alternatively, it can be proved from the sum rule together with the constant multiple rule:

$$\begin{aligned}(f-g)'(x) &= (f+(-1)g)'(x) \\ &= f'(x) + ((-1)g)'(x) && \text{Sum Rule} \\ &= f'(x) + (-1)\cdot g'(x) && \text{Constant Multiple Rule} \\ &= f'(x) - g'(x).\end{aligned}$$

- Term by Term Differentiation can be applied to sum and difference of finitely many terms.

Example Let $y = x^2 + 3$. Find $\frac{dy}{dx}$.

Solution
$$\begin{aligned}\frac{dy}{dx} &= \frac{d}{dx}(x^2+3) \\ &= \frac{d}{dx}x^2 + \frac{d}{dx}3 && \text{Term by Term Differentiation} \\ &= 2x + 0 && \text{Power Rule and Derivative of Constant} \\ &= 2x\end{aligned}$$
□

Example Let $f(x) = x^5 - 6x^7$. Find $f'(x)$.

Explanation This question is similar to the last one. If we put $y = f(x)$, then $f'(x) = \frac{dy}{dx}$. Below, we use the notation $\frac{d}{dx}$ to perform differentiation.

Solution
$$\begin{aligned}f'(x) &= \frac{d}{dx}(x^5 - 6x^7) \\ &= \frac{d}{dx}x^5 - \frac{d}{dx}6x^7 && \text{Term by Term Differentiation} \\ &= 5x^4 - 6\cdot\frac{d}{dx}x^7 && \text{Power Rule and Constant Multiple Rule} \\ &= 5x^4 - 6\cdot(7x^6) && \text{Power Rule} \\ &= 5x^4 - 42x^6\end{aligned}$$
□

The last two examples illustrate that polynomials can be differentiated term by term.

Derivative of Polynomial *Let* $y = a_n x^n + a_{n-1}x^{n-1} + \cdots + a_2x^2 + a_1x + a_0$ *be a polynomial. Then we have*

$$\frac{\mathrm{d}y}{\mathrm{d}x} = na_nx^{n-1} + (n-1)a_{n-1}x^{n-2} + \cdots + 2a_2x + a_1.$$

Proof

$$\begin{aligned}
\frac{\mathrm{d}y}{\mathrm{d}x} &= \frac{\mathrm{d}}{\mathrm{d}x}\left(a_nx^n + a_{n-1}x^{n-1} + \cdots + a_2x^2 + a_1x + a_0\right) \\
&= \frac{\mathrm{d}}{\mathrm{d}x}a_nx^n + \frac{\mathrm{d}}{\mathrm{d}x}a_{n-1}x^{n-1} + \cdots + \frac{\mathrm{d}}{\mathrm{d}x}a_2x^2 + \frac{\mathrm{d}}{\mathrm{d}x}a_1x + \frac{\mathrm{d}}{\mathrm{d}x}a_0 && \text{Term by Term Differentiation} \\
&= a_n\frac{\mathrm{d}}{\mathrm{d}x}x^n + a_{n-1}\frac{\mathrm{d}}{\mathrm{d}x}x^{n-1} + \cdots + a_2\frac{\mathrm{d}}{\mathrm{d}x}x^2 + a_1\frac{\mathrm{d}}{\mathrm{d}x}x + 0 && \text{Constant Multiple Rule and Derivative of Constant} \\
&= a_nnx^{n-1} + a_{n-1}(n-1)x^{n-2} + \cdots + a_2\cdot(2x) + a_1\cdot 1 && \text{Power Rule}
\end{aligned}$$

□

Example Let $f(x) = 2x(x^2 - 5x + 7)$. Find the derivative of f at 2.

Explanation The question is to find $f'(2)$. We can find $f'(x)$ using one of the following two ways:

- product rule (will be discussed later);
- expand the expression to get a polynomial.

Then putting $x = 2$, we get the answer. Below we use the second method to find $f'(x)$.

Solution

$$\begin{aligned}
f'(x) &= \frac{\mathrm{d}}{\mathrm{d}x}\left(2x(x^2 - 5x + 7)\right) \\
&= \frac{\mathrm{d}}{\mathrm{d}x}\left(2x^3 - 10x^2 + 14x\right) \\
&= 2\cdot(3x^2) - 10\cdot(2x) + 14\cdot 1 && \text{Derivative of Polynomial} \\
&= 6x^2 - 20x + 14
\end{aligned}$$

The derivative of f at 2 is

$$\begin{aligned}
f'(2) &= 6(2^2) - 20(2) + 14 \\
&= -2
\end{aligned}$$

□

Product Rule *Let* f *and* g *be functions with the same domain. Suppose that* f *and* g *are differentiable. Then the function* fg *is also differentiable. Moreover, we have*

$$\frac{\mathrm{d}}{\mathrm{d}x}(fg)(x) = g(x)\cdot\frac{\mathrm{d}}{\mathrm{d}x}f(x) + f(x)\cdot\frac{\mathrm{d}}{\mathrm{d}x}g(x).$$

Explanation The function fg is defined by $(fg)(x) = f(x)\cdot g(x)$ for x belonging to the common domain A of f and g. The result means that if $f'(x)$ and $g'(x)$ exist for all $x \in A$, then $(fg)'(x) = g(x)f'(x) + f(x)g'(x)$ for all $x \in A$, that is, $(fg)' = gf' + fg'$.

Proof By definition, we have

$$\begin{aligned}
(fg)'(x) &= \lim_{h\to 0}\frac{(fg)(x+h) - (fg)(x)}{h} \\
&= \lim_{h\to 0}\frac{f(x+h)g(x+h) - f(x)g(x)}{h} && \text{Definition of } fg
\end{aligned}$$

To find the limit, we use the following technique: "*subtract and add*" $f(x+h)g(x)$ in the numerator.

$$
\begin{aligned}
(fg)'(x) &= \lim_{h\to 0}\frac{f(x+h)g(x+h)-f(x+h)g(x)+f(x+h)g(x)-f(x)g(x)}{h} \\
&= \lim_{h\to 0}\left(\frac{f(x+h)g(x+h)-f(x+h)g(x)}{h}+\frac{f(x+h)g(x)-f(x)g(x)}{h}\right) \\
&= \lim_{h\to 0}\left(f(x+h)\cdot\frac{g(x+h)-g(x)}{h}\right)+\lim_{h\to 0}\left(g(x)\cdot\frac{f(x+h)-f(x)}{h}\right) && \text{Limit Rule (La4)} \\
&= \lim_{h\to 0}f(x+h)\cdot\lim_{h\to 0}\frac{g(x+h)-g(x)}{h}+\lim_{h\to 0}g(x)\cdot\lim_{h\to 0}\frac{f(x+h)-f(x)}{h} && \text{Limit Rule (La5)} \\
&= f(x)g'(x)+g(x)f'(x) && \text{Theorem 4.1.1 \& Limit Rule (La1)}
\end{aligned}
$$

In the last step, the first limit is found by substitution because f is continuous; the third limit is $g(x)$ because considered as a function of h, $g(x)$ is a constant. □

Example Let $y=(x+1)(x^2+3)$. Find $\dfrac{dy}{dx}$.

Explanation The expression defining y is a product of two functions. So we can apply the product rule. Alternatively, we can expand the expression to get a polynomial and then differentiate term by term.

Solution 1

$$
\begin{aligned}
\frac{dy}{dx} &= \frac{d}{dx}\big((x+1)(x^2+3)\big) \\
&= (x^2+3)\cdot\frac{d}{dx}(x+1)+(x+1)\cdot\frac{d}{dx}(x^2+3) && \text{Product Rule} \\
&= (x^2+3)\cdot(1+0)+(x+1)\cdot(2x+0) && \text{Derivative of Polynomial} \\
&= 3x^2+2x+3
\end{aligned}
$$

□

Solution 2

$$
\begin{aligned}
\frac{dy}{dx} &= \frac{d}{dx}\big((x+1)(x^2+3)\big) \\
&= \frac{d}{dx}(x^3+x^2+3x+3) \\
&= 3x^2+2x+3 && \text{Derivative of Polynomial}
\end{aligned}
$$

□

Quotient Rule Let f and g be functions with the same domain. Suppose that f and g are differentiable and that g has no zero in its domain. Then the function $\frac{f}{g}$ is also differentiable. Moreover, we have,

$$\frac{d}{dx}\left(\frac{f}{g}\right)(x)=\frac{g(x)\cdot\frac{d}{dx}f(x)-f(x)\cdot\frac{d}{dx}g(x)}{(g(x))^2}.$$

Explanation The condition "*g has no zero in its domain*" means that $g(x)\neq 0$ for all $x\in \operatorname{dom}(g)$. The function $\frac{f}{g}$ is defined by $\left(\frac{f}{g}\right)(x)=\frac{f(x)}{g(x)}$ for $x\in A$, where A is the common domain of f and g. The result means that if $f'(x)$ and $g'(x)$ exist for all $x\in A$, then $\left(\frac{f}{g}\right)'(x)=\frac{g(x)f'(x)-f(x)g'(x)}{g(x)^2}$ for all $x\in A$, that is, $\left(\frac{f}{g}\right)'=\frac{gf'-fg'}{g^2}$.

Proof The proof is similar to that for the product rule. □

Example Let $y = \dfrac{x^2+3x-4}{2x+1}$. Find $\dfrac{dy}{dx}$.

Solution

$$\begin{aligned}
\frac{dy}{dx} &= \frac{d}{dx}\left(\frac{x^2+3x-4}{2x+1}\right) \\
&= \frac{(2x+1)\cdot\frac{d}{dx}(x^2+3x-4)-(x^2+3x-4)\cdot\frac{d}{dx}(2x+1)}{(2x+1)^2} && \text{Quotient Rule} \\
&= \frac{(2x+1)(2x+3)-(x^2+3x-4)(2)}{(2x+1)^2} && \text{Derivative of Polynomial} \\
&= \frac{(4x^2+8x+3)-(2x^2+6x-8)}{(2x+1)^2} \\
&= \frac{2x^2+2x+11}{(2x+1)^2}
\end{aligned}$$

□

Power Rule for Differentiation (negative integer version) *Let n be a negative integer. Then the power function x^n is differentiable on $\mathbb{R}\setminus\{0\}$ and we have*

$$\frac{d}{dx}x^n = nx^{n-1}.$$

Explanation Since n is a negative integer, it can be written as $-m$ where m is a positive integer. The function $x^n = x^{-m} = \frac{1}{x^m}$ is defined for all $x \neq 0$, that is, the domain of x^n is $\mathbb{R}\setminus\{0\}$.

Proof Let $f : \mathbb{R}\setminus\{0\} \longrightarrow \mathbb{R}$ be the function given by $f(x) = x^{-m}$, where $m = -n$. By definition, we have

$$\begin{aligned}
f'(x) &= \frac{d}{dx}\frac{1}{x^m} \\
&= \frac{x^m\cdot\frac{d}{dx}1 - 1\cdot\frac{d}{dx}x^m}{(x^m)^2} && \text{Quotient Rule} \\
&= \frac{x^m\cdot 0 - 1\cdot mx^{m-1}}{x^{2m}} && \text{Derivative of Constant \& Power Rule (positive integer version)} \\
&= -mx^{(m-1)-2m} \\
&= -mx^{-m-1} \\
&= nx^{n-1}
\end{aligned}$$

□

Example Find an equation for the tangent line to the curve $y = \frac{3x^2-1}{x}$ at the point $(1,2)$.

Explanation The curve is given by $y = f(x)$ where $f(x) = \frac{3x^2-1}{x}$. Since $f(1) = 2$, the point $(1,2)$ lies on the curve. To find an equation for the tangent line, we have to find the slope at the point (and then use point-slope form). The slope at the point $(1,2)$ is $f'(1)$. We can use rules for differentiation to find $f'(x)$ and then substitute $x = 1$ to get $f'(1)$.

Solution To find $\frac{dy}{dx}$, we can use quotient rule or term by term differentiation.

$$(\textit{Method 1}) \quad \begin{aligned} \frac{\mathrm{d}y}{\mathrm{d}x} &= \frac{\mathrm{d}}{\mathrm{d}x}\frac{3x^2-1}{x} \\ &= \frac{x \cdot \frac{\mathrm{d}}{\mathrm{d}x}(3x^2-1) - (3x^2-1)\cdot\frac{\mathrm{d}}{\mathrm{d}x}x}{x^2} && \text{Quotient Rule} \\ &= \frac{x\cdot 3\cdot(2x) - (3x^2-1)\cdot 1}{x^2} && \text{Derivative of Polynomial} \\ &= \frac{3x^2+1}{x^2} \end{aligned}$$

$$(\textit{Method 2}) \quad \begin{aligned} \frac{\mathrm{d}y}{\mathrm{d}x} &= \frac{\mathrm{d}}{\mathrm{d}x}\left((3x^2-1)x^{-1}\right) \\ &= \frac{\mathrm{d}}{\mathrm{d}x}(3x - x^{-1}) \\ &= \frac{\mathrm{d}}{\mathrm{d}x}3x - \frac{\mathrm{d}}{\mathrm{d}x}x^{-1} && \text{Term by Term Differentiation} \\ &= 3 - (-1)x^{-2} && \text{Derivative of Polynomial \& Power Rule} \\ &= 3 + \frac{1}{x^2}. \end{aligned}$$

The slope of the tangent line at the point $(1, 2)$ is $\left.\frac{\mathrm{d}y}{\mathrm{d}x}\right|_{x=1} = 3 + \frac{1}{1} = 4.$

Equation for the tangent line: $y - 2 = 4(x - 1)$ (point-slope form)

$4x - y - 2 = 0$ (general linear form) □

The next result is a special case of the general power rule. Since the square root function appears often in applied problems and the proof is not difficult, we give the result here.

Derivative of the Square Root Function The derivative of the square root function $\sqrt{x}$ is $\frac{1}{2\sqrt{x}}$, that is,

$$\frac{\mathrm{d}}{\mathrm{d}x}\sqrt{x} = \frac{1}{2\sqrt{x}}.$$

Explanation The domain of the square root function is $[0, \infty)$. Since the function is undefined on the left-side of 0, we can only consider differentiability of the function on $(0, \infty)$. The result means that $f'(x) = \frac{1}{2\sqrt{x}}$ for all $x \in (0, \infty)$ where $f(x) = \sqrt{x}$.

Proof Let $f : [0, \infty) \longrightarrow \mathbb{R}$ be the function given by $f(x) = \sqrt{x}$. For every $x > 0$, note that if h is a small enough real number, then $x + h \in \operatorname{dom}(f)$ and hence if h is a small enough non-zero real number, we have

$$\begin{aligned} \frac{f(x+h) - f(x)}{h} &= \frac{\sqrt{x+h} - \sqrt{x}}{h} = \frac{\left(\sqrt{x+h} - \sqrt{x}\right)\left(\sqrt{x+h} + \sqrt{x}\right)}{h\left(\sqrt{x+h} + \sqrt{x}\right)} \\ &= \frac{(x+h) - x}{h\left(\sqrt{x+h} + \sqrt{x}\right)} = \frac{h}{h\left(\sqrt{x+h} + \sqrt{x}\right)} \\ &= \frac{1}{\sqrt{x+h} + \sqrt{x}} \end{aligned}$$

which, by definition, implies that

$$\begin{aligned}
f'(x) &= \lim_{h\to 0} \frac{1}{\sqrt{x+h}+\sqrt{x}} \\
&= \frac{1}{\lim_{h\to 0}\left(\sqrt{x+h}+\sqrt{x}\right)} && \text{Limit Rule (La6)} \\
&= \frac{1}{\lim_{h\to 0}\sqrt{x+h}+\lim_{h\to 0}\sqrt{x}} && \text{Limit Rule (La4)} \\
&= \frac{1}{\sqrt{x+0}+\sqrt{x}} && \text{By continuity \& Limit Rule (La1)} \\
&= \frac{1}{2\sqrt{x}}
\end{aligned}$$

In the second last step, the first limit is found by substitution because the square root function is continuous; the second limit is the limit of a constant. □

Example Let $y = \sqrt{x}\,(x+1)$. Find $\dfrac{dy}{dx}$.

Solution

$$\begin{aligned}
\frac{dy}{dx} &= \frac{d}{dx}\sqrt{x}\,(x+1) \\
&= (x+1)\cdot\frac{d}{dx}\sqrt{x}+\sqrt{x}\cdot\frac{d}{dx}(x+1) && \text{Product Rule} \\
&= (x+1)\cdot\frac{1}{2\sqrt{x}}+\sqrt{x}\cdot 1 && \text{Derivative of Square Root Function \& Derivative of Polynomial} \\
&= \frac{\sqrt{x}}{2}+\frac{1}{2\sqrt{x}}+\sqrt{x} \\
&= \frac{3\sqrt{x}}{2}+\frac{1}{2\sqrt{x}}
\end{aligned}$$

□

Remark If we expand the expression defining y, we get $x^{\frac{3}{2}} + x^{\frac{1}{2}}$. The derivative can be found if we know the derivative of $x^{\frac{3}{2}}$.

Power Rule for Differentiation ($n + \frac{1}{2}$ version) *Let n be an integer. Then the function $x^{n+\frac{1}{2}}$ is differentiable on $(0, \infty)$ and we have*

$$\frac{d}{dx}x^{n+\frac{1}{2}} = \left(n+\frac{1}{2}\right)x^{n-\frac{1}{2}}.$$

Explanation Denoting the function by f, if $n = 0$, then f is the square root function; if n is positive, then the domain of f is $[0, \infty)$; if n is negative, then the domain of f is $(0, \infty)$. Putting $r = n + \frac{1}{2}$, then we have $f(x) = x^r$ and the result means that $f'(x) = rx^{r-1}$ for all $x > 0$. This result has the same form as the power rule and it will be referred to as the *power rule*.

Remark There is a more general result called the *General Power Rule* (see Chapter 9).

Proof Let f be the function given by $f(x) = x^{n+\frac{1}{2}}$. Note that $f(x) = x^n \cdot \sqrt{x}$ and so by the Product Rule and the

Rule for Derivative of the Square Root Function, the function f is differentiable on $(0, \infty)$ and we have

$$\begin{aligned}
f'(x) &= \frac{d}{dx}\left(x^n \cdot \sqrt{x}\right) \\
&= \sqrt{x} \cdot \frac{d}{dx}x^n + x^n \cdot \frac{d}{dx}\sqrt{x} && \text{Product Rule} \\
&= \sqrt{x} \cdot nx^{n-1} + x^n \cdot \frac{1}{2\sqrt{x}} && \text{Power Rule \& Derivative Square Root Function} \\
&= nx^{n-\frac{1}{2}} + \frac{1}{2}x^{n-\frac{1}{2}} \\
&= \left(n + \frac{1}{2}\right)x^{n-\frac{1}{2}}
\end{aligned}$$

□

Below, we redo the preceding example using the $n + \frac{1}{2}$ version of the power rule.

Example Let $y = \sqrt{x}\,(x + 1)$. Find $\frac{dy}{dx}$.

Solution

$$\begin{aligned}
\frac{dy}{dx} &= \frac{d}{dx}x^{\frac{1}{2}}(x + 1) \\
&= \frac{d}{dx}\left(x^{\frac{3}{2}} + x^{\frac{1}{2}}\right) \\
&= \frac{d}{dx}x^{\frac{3}{2}} + \frac{d}{dx}x^{\frac{1}{2}} && \text{Term by Term Differentiation} \\
&= \frac{3}{2}x^{\frac{3}{2}-1} + \frac{1}{2}x^{\frac{1}{2}-1} && \text{Power Rule} \\
&= \frac{3}{2}x^{\frac{1}{2}} + \frac{1}{2}x^{-\frac{1}{2}}
\end{aligned}$$

□

Example Let $y = \frac{2x^2 - 3}{\sqrt{x}}$. Find $\frac{dy}{dx}$.

Solution 1

$$\begin{aligned}
\frac{dy}{dx} &= \frac{d}{dx}\frac{2x^2 - 3}{\sqrt{x}} \\
&= \frac{\sqrt{x} \cdot \frac{d}{dx}(2x^2 - 3) - (2x^2 - 3) \cdot \frac{d}{dx}\sqrt{x}}{(\sqrt{x})^2} && \text{Quotient Rule} \\
&= \frac{\sqrt{x} \cdot 2 \cdot (2x) - (2x^2 - 3) \cdot \frac{1}{2\sqrt{x}}}{x} && \text{Derivative of Polynomial \& Derivative of Square Root Function} \\
&= \frac{4x\sqrt{x} - x\sqrt{x} + \frac{3}{2\sqrt{x}}}{x} \\
&= 3\sqrt{x} + \frac{3}{2x\sqrt{x}}
\end{aligned}$$

□

Solution 2
$$\begin{aligned}\frac{dy}{dx} &= \frac{d}{dx}(2x^2-3)x^{-\frac{1}{2}} \\ &= \frac{d}{dx}\left(2x^{\frac{3}{2}}-3x^{-\frac{1}{2}}\right) \\ &= 2\cdot\frac{d}{dx}x^{\frac{3}{2}}-3\cdot\frac{d}{dx}x^{-\frac{1}{2}} && \text{Term by Term Differentiation} \\ &= 2\cdot\left(\frac{3}{2}x^{\frac{1}{2}}\right)-3\cdot\left(-\frac{1}{2}x^{-\frac{3}{2}}\right) && \text{Power Rule} \\ &= 3x^{\frac{1}{2}}+\frac{3}{2}x^{-\frac{3}{2}}\end{aligned}$$
□

We close this section with the following "*Caution*" and "*Question*". The "*caution*" points out a common mistake that many students made.

Caution $\frac{d}{dx}2^x \neq x\cdot 2^{x-1}$, we can't apply the Power Rule because 2^x is not a power function; it is an exponential function (see Chapter 8).

Question Using rules discussed in this chapter, we can differentiate polynomial functions as well as rational functions. For example, to differentiate $f(x)=(x^2+5)^3$, we can first expand the cube to get a polynomial of degree 6 and then differentiate term by term. How about the following

(1) $f(x)=(x^2+5)^{30}$?

(2) $f(x)=\sqrt{x^2+5}$?

For (1), we can expand and then differentiate term by term (if we have enough patience). However, this doesn't work for (2). Note that both functions are in the form $x\mapsto(x^2+5)^r$ which can be considered as composition of two functions:

$$x\mapsto(x^2+5)\mapsto(x^2+5)^r.$$

In Chapter 9, we will discuss the *chain rule*, a tool to handle this kind of differentiation.

Exercise 4.2

1. For each of the following y, find $\frac{dy}{dx}$.

   (a) $y=-\pi$ (b) $y=2x^9+3x$

   (c) $y=x^2+5x-7$ (d) $y=x(x-1)$

   (e) $y=(2x-3)(5-6x)$ (f) $y=(x^2+5)^3$

   (g) $y=\frac{23}{x^4}$ (h) $y=\frac{x-1}{x}$

   (i) $y=\frac{x-1}{x+1}$ (j) $y=\sqrt{x}(\sqrt{x}+1)$

2. For each of the following f, find $f'(a)$ for the given a.

   (a) $f(x)=x^3-4x,\quad a=1$ (b) $f(x)=\frac{2}{x^3}+\frac{4}{x},\quad a=2$

   (c) $f(x)=3x^{\frac{1}{3}}-5x^{\frac{2}{3}},\quad a=27$ (d) $f(x)=\pi x^2-2\sqrt{x},\quad a=4$

   (e) $f(x)=(x^2+3)(x^3+2),\quad a=1$ (f) $f(x)=\frac{x^2+1}{2x-3},\quad a=2$

   (g) $f(x)=\frac{x^2+3x-5}{x^2-7x+5},\quad a=1$

3. Consider the curve given by $y = 3x^4 - 6x^2 + 2$.
   (a) Find the slope of the curve at the point whose x-coordinate is 2.
   (b) Find an equation for the tangent line at the point $(1, -1)$.
   (c) Find the point(s) on the curve at which the tangent line is horizontal.

4. *It is known that* $\frac{\mathrm{d}}{\mathrm{d}x} \sin x = \cos x$. For each of the following y, find $\frac{\mathrm{d}y}{\mathrm{d}x}$.
   (a) $y = x \sin x$ (b) $y = \frac{\sin x}{x+1}$
   (c) $y = \frac{x^2+1}{\sin x}$ (d) $y = x(x+2)\sin x$

5. Let f be a differentiable function.
   (a) Use product rule to show that $\frac{\mathrm{d}}{\mathrm{d}x} f(x)^2 = 2f(x) \cdot \frac{\mathrm{d}}{\mathrm{d}x} f(x)$.
   (b) Use product rule and the result in (a) to show that $\frac{\mathrm{d}}{\mathrm{d}x} f(x)^3 = 3f(x)^2 \cdot \frac{\mathrm{d}}{\mathrm{d}x} f(x)$.
   (c) Can you guess (and prove) a formula for $\frac{\mathrm{d}}{\mathrm{d}x} f(x)^n$, where n is a positive integer?

6. (a) Use the result in 5(a) to find $\frac{\mathrm{d}}{\mathrm{d}x}(x^3 + 5x^2 - 2)^2$.
   (b) Use the result in 5(b) to find $\frac{\mathrm{d}}{\mathrm{d}x}(x^2 + 5)^3$. *Compare your answer with that in Q.1(f).*

4.3 Higher-Order Derivatives

Let f be a function that is differentiable at some points belonging to $\mathrm{dom}(f)$. Then f' is a function.

- If, in addition, f' is differentiable at some points belonging to $\mathrm{dom}(f')$, then the derivative of f' exists and is denoted by f''; it is the function given by $f''(x) = \lim_{h \to 0} \frac{f'(x+h) - f'(x)}{h}$ and is called the *second derivative* of f.
- If, in addition, f'' is differentiable at some points belonging to $\mathrm{dom}(f'')$, then the derivative of f'' is denoted by f''', called the *third derivative* of f.
- In general, the n-th derivative of f (where n is a positive integer), denoted by $f^{(n)}$, is defined to be the derivative of the $(n-1)$-th derivative of f (where the 0-th derivative of f means f). For $n = 1$, the first derivative of f is simply the derivative f' of f. For $n > 1$, $f^{(n)}$ is called a *higher-order derivative* of f.

Notation Similar to first order derivative, we have different notations for second order derivative of f.

$$y'', \quad f'', \quad \frac{\mathrm{d}^2 y}{\mathrm{d}x^2}, \quad D^2 f, \quad D^2 y, \quad f''(x) \quad \text{and} \quad \frac{\mathrm{d}^2}{\mathrm{d}x^2} f(x).$$

Readers may compare these with that on page 109. Similarly, we also have different notations for other higher-order derivatives.

Example Let $f(x) = 5x^3 - 2x^2 + 6x + 1$. Find the derivative and all the higher-order derivatives of f.

Explanation The question is to find for each positive integer n, the domain of the n-th derivative of f and a formula for $f^{(n)}(x)$. To find $f'(x)$, we can apply differentiation term by term. To find $f''(x)$, by definition, we have $f''(x) = \frac{\mathrm{d}}{\mathrm{d}x} f'(x)$ which can be simplified using the result for $f'(x)$ and rules for differentiation.

Solution

$$\begin{aligned}
f'(x) &= \frac{\mathrm{d}}{\mathrm{d}x}(5x^3 - 2x^2 + 6x + 1) \\
&= 15x^2 - 4x + 6 && \text{Derivative of Polynomial} \\
f''(x) &= \frac{\mathrm{d}}{\mathrm{d}x}(15x^2 - 4x + 6) \\
&= 30x - 4 && \text{Derivative of Polynomial} \\
f'''(x) &= \frac{\mathrm{d}}{\mathrm{d}x}(30x - 4) \\
&= 30 && \text{Derivative of Polynomial} \\
f^{(4)}(x) &= 0 && \text{Derivative of Constant}
\end{aligned}$$

From this we see that for $n \geq 4$, $f^{(n)}(x) = 0$. Moreover, for every positive integer n, the domain of $f^{(n)}$ is $\mathbb{R}$. □

Example Let $f(x) = \frac{x^3 - 1}{x}$. Find $f'(3)$ and $f''(-4)$.

Explanation

- To find $f'(3)$, we find $f'(x)$ first and then substitute $x = 3$. Although $f(x)$ is written as a quotient of two functions, it is better to find $f'(x)$ by expanding $(x^3 - 1)x^{-1}$.
- To find $f''(-4)$, we find $f''(x)$ first and then substitute $x = -4$. To find $f''(x)$, we differentiate the result obtained for $f'(x)$.

Solution

$$\begin{aligned}
f'(x) &= \frac{\mathrm{d}}{\mathrm{d}x}\left((x^3 - 1)x^{-1}\right) \\
&= \frac{\mathrm{d}}{\mathrm{d}x}\left(x^2 - x^{-1}\right) \\
&= 2x - (-1)x^{-2} && \text{Term by Term Differentiation \& Power Rule} \\
&= 2x + x^{-2} \\
f'(3) &= 2 \cdot (3) + 3^{-2} \\
&= \frac{55}{9} \\
f''(x) &= \frac{\mathrm{d}}{\mathrm{d}x}\left(2x + x^{-2}\right) && \text{By result for } f'(x) \\
&= 2 + (-2)x^{-3} && \text{Term by Term Differentiation \& Power Rule} \\
&= 2 - 2x^{-3} \\
f''(-4) &= 2 - 2 \cdot (-4)^{-3} \\
&= \frac{65}{32}
\end{aligned}$$

□

Meaning of Second Derivative

- The graph of $y = f(x)$ is a curve. Note that $f'(x) = \frac{\mathrm{d}y}{\mathrm{d}x}$ is the slope function; it is the rate of change of y with respect to x. Since $f''(x) = \frac{\mathrm{d}^2y}{\mathrm{d}x^2}$ is the derivative of the slope function, it is the rate of change

of slope and is related to a concept called *convexity* (*bending*) of a curve. More details can be found in Chapter 5.

- If $x = t$ is time and if $y = s(t)$ is the displacement function of a moving object, then $s'(t) = \frac{\mathrm{d}s}{\mathrm{d}t}$ is the velocity function. The derivative of velocity is $s''(t)$ or $\frac{\mathrm{d}^2 s}{\mathrm{d}t^2}$; it is the rate of change of the velocity (function), that is, the *acceleration* (function).

Exercise 4.3

1. For each of the following y, find $\frac{\mathrm{d}^2 y}{\mathrm{d}x^2}$.

   (a) $y = x^3 - 3x^2 + 4x - 1$ (b) $y = (2x + 3)(4 - x)$

   (c) $y = \sqrt{x}\,(1 + x)$ (d) $y = \frac{1 - 2x}{x^2}$

   (e) $y = (x^3 + 1)^2$

2. For each of the following f, find $f''(a)$ for the given a.

   (a) $f(x) = 7x^6 - 8x^5 + 15x, \qquad a = 1$

   (b) $f(x) = x^2(1 - 2x), \qquad a = 2$

   (c) $f(x) = (2 + 3x)^2, \qquad a = 0$

3. Let $f(x) = a_n x^n + a_{n-1} x^{n-1} + \cdots + a_1 x + a_0$ be a polynomial of degree n.

   (a) Find $f(0)$ and $f'(0)$.

   (b) Find $f^{(n)}(x)$ and $f^{(n+1)}(x)$.

Chapter 5

Applications of Differentiation

In this chapter, we will discuss applications of differentiation to curve sketching and extremal problems. For curve sketching, we need to consider geometric meanings of the first and second order derivatives. For convenience, some of the concepts and results are given not in their most general forms. Many of the concepts and results are stated for functions that are differentiable, twice differentiable etc. Below are the meanings of these terms.

Terminology Let f be a function that is defined on an open interval (a, b). We say that

- f is *differentiable* on (a, b) if $f'(x)$ exists for all $x \in (a, b)$;
- f is *twice differentiable* on (a, b) if $f''(x)$ exists for all $x \in (a, b)$.

Explanation The condition "*f is a function defined on an open interval* (a, b)" means that $(a, b) \subseteq \operatorname{dom}(f)$.

Terminology Let f be a function and let x_0 be a real number. We say that

- f is *defined on an open interval containing* x_0 if there is an open interval (a, b) such that $x_0 \in (a, b)$ and $(a, b) \subseteq \operatorname{dom}(f)$;
- f is *differentiable on an open interval containing* x_0 if there is an open interval (a, b) such that $x_0 \in (a, b)$ and f is differentiable on (a, b);

Remark If f is a function defined on an open interval containing x_0, then we may consider continuity and differentiability of f at x_0.

Example Let $f(x) = \sqrt{x}$. The domain of f is $[0, \infty)$.

- Although $0 \in \operatorname{dom}(f)$, the function f is not defined on an open interval containing 0. This is because there does not exist any open interval (a, b) such that $0 \in (a, b)$ and $(a, b) \subseteq \operatorname{dom}(f)$.
- If x_0 is a positive real number, then f is defined on an open interval containing x_0. This is because $x_0 \in (0, \infty)$ and $(0, \infty) \subseteq \operatorname{dom}(f)$.

For $x > 0$, by the Power Rule, we have $f'(x) = \dfrac{1}{2\sqrt{x}}$. Thus, f is differentiable on $(0, \infty)$.

- If x_0 is a positive real number, then f is differentiable on an open interval containing x_0.

For $x > 0$, by the Constant Multiple Rule and the Power Rule, we have $f''(x) = \dfrac{-1}{4\sqrt{x^3}}$. Thus, f is twice differentiable on $(0, \infty)$.

5.1 Curve Sketching

5.1.1 Increasing and Decreasing Functions

Definition Let f be a function and let I be an interval such that $I \subseteq \text{dom}\,(f)$. We say that f is

- *strictly increasing* on I if for any two numbers $x_1, x_2 \in I$, where $x_1 < x_2$, we have $f(x_1) < f(x_2)$;
- *strictly decreasing* on I if for any two numbers x_1, x_2 in I, where $x_1 < x_2$, we have $f(x_1) > f(x_2)$.

Remark

- In the definition, I can be an open interval, a closed interval or a half-open half-closed interval in the form $[a, b)$ or $(a, b]$.
- Although we can define the concepts "*f is strictly increasing (or strictly decreasing) on a set S, where* $S \subseteq \text{dom}\,(f)$", we will not use such concepts in this course because the concepts "*strictly increasing (or strictly decreasing) on an interval*" are good enough for our consideration; moreover, a function strictly increasing (or strictly decreasing) on a set S_1 and also on a set S_2 may not be strictly increasing (or strictly decreasing) on $S_1 \cup S_2$.

Geometric Meaning A function is strictly increasing (respectively strictly decreasing) on an interval I means that for $x \in I$, the graph of f goes up (respectively goes down) as x goes from left to right.

Terminology For simplicity, instead of saying "*strictly increasing*", we will say "*increasing*" etc.

Remark Some authors have a different definition for "*increasing*". In that case, "*increasing*" and "*strictly increasing*" refer to different concepts.

Example Let $f : \mathbb{R} \longrightarrow \mathbb{R}$ be the function given by $f(x) = x^3$.
Then f is increasing on $\mathbb{R}$.

Reason If $x_1 < x_2$, then $x_1^3 < x_2^3$.

Figure 5.1

Example Let $f : (0, \infty) \longrightarrow \mathbb{R}$ be the function given by $f(x) = \dfrac{1}{\sqrt{x}}$.
Then f is decreasing on $(0, \infty)$.

Reason If $0 < x_1 < x_2$, then $\sqrt{x_1} < \sqrt{x_2}$ and so $\dfrac{1}{\sqrt{x_1}} > \dfrac{1}{\sqrt{x_2}}$.

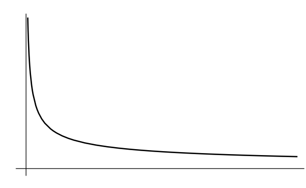


Figure 5.2

Example Let $f : \mathbb{R} \longrightarrow \mathbb{R}$ be the function given by $f(x) = x^2$.
Then f is decreasing on $(-\infty, 0)$ and increasing on $(0, \infty)$.

Reason

- If $x_1 < x_2 < 0$, then $x_1^2 > x_2^2$.
- If $0 < x_1 < x_2$, then $x_1^2 < x_2^2$.

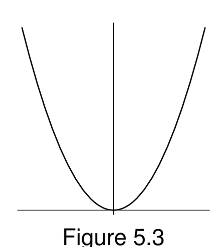


Figure 5.3

Remark

- If a function is increasing (respectively decreasing) on an interval I, then it is also increasing (respectively decreasing) on any interval J with $J \subseteq I$. For the function f in the above example, we can also say that f is increasing on $(0, 1]$; f is decreasing on $[-10, -2]$ etc.
- If a function is continuous on an interval $[a, b)$ and if it is increasing (respectively decreasing) on the open interval (a, b), then it is increasing (respectively decreasing) on $[a, b)$. Similar results holds if $[a, b)$ is replaced by $(a, b]$ or $[a, b]$. For the function in the above example, it is decreasing on $(-\infty, 0]$ and increasing on $[0, \infty)$. These intervals are maximal in the sense that they cannot be enlarged.

Definition Let f be a function and let I be an interval with $I \subseteq \text{dom}(f)$ such that f is increasing (respectively decreasing) on I. We say that I is a *maximal interval* on which f is increasing (respectively decreasing) if there does not exist any interval J with $I \subsetneq J \subseteq \text{dom}(f)$ such that f is increasing (respectively decreasing) on J.

Example For the function f given in the preceding example, the interval $(-\infty, 0]$ is the maximal interval on which f is decreasing and $[0, \infty)$ is the maximal interval on which f is increasing.

In the above examples, we can determine where the function f is increasing or decreasing using inspection or using the graph of f. In general, given a function f, for example, $f(x) = 27x - x^3$, it is not easy to see where f is increasing or decreasing. For differentiable functions, the next theorem describes a simple way to determine where f is increasing or decreasing.

Theorem 5.1.1 *Let f be a function that is defined and differentiable on an open interval (a, b).*

(1) *If $f'(x) > 0$ for all $x \in (a, b)$, then f is increasing on (a, b).*

(2) *If $f'(x) < 0$ for all $x \in (a, b)$, then f is decreasing on (a, b).*

(3) *If $f'(x) = 0$ for all $x \in (a, b)$, then f is constant on (a, b), that is, $f(x_1) = f(x_2)$ for all $x_1, x_2 \in (a, b)$, or equivalently, there exists a real number c such that $f(x) = c$ for all $x \in (a, b)$.*

Explanation In the first result, the condition "$f'(x) > 0$ *for all* $x \in (a, b)$" means that the slope is always positive. From intuition, we "*see*" that the graph of f goes up. However, this is not a proof.

Proof The results can be proved rigorously using a result called the Mean-Value Theorem. For details, see Theorem B.3.1 in the Appendix. □

Example Let $f : \mathbb{R} \longrightarrow \mathbb{R}$ be the function given by

$$f(x) = 27x - x^3.$$

Find the interval(s), if any, on which f is increasing or decreasing.

Explanation

- The question is to find maximal interval(s) on which f is increasing and to find maximal interval(s) on which f is decreasing.

- Before getting the answer, we do not know whether there is any interval on which f is increasing or decreasing, so in the question, "*if any*" is added. Moreover, we do not know whether there are more than one intervals on which f is increasing or decreasing, so instead of asking for "*interval*", we ask for "*interval(s)*". This kind of wording is cumbersome; therefore, sometimes, we simply ask: "*Find the intervals on which f is increasing or decreasing*".
- Because f is continuous, it suffices to find maximal open intervals on which f is increasing or decreasing. For this, we have to solve $f'(x) > 0$ or $f'(x) < 0$ respectively.

Solution Differentiating $f(x)$, we get

$$\begin{aligned} f'(x) &= \frac{\mathrm{d}}{\mathrm{d}x}(27x - x^3) \\ &= 27 - 3x^2 \\ &= 3(3 + x)(3 - x). \end{aligned}$$

	$(-\infty, -3)$	$(-3, 3)$	$(3, \infty)$
3	+	+	+
$3 + x$	−	+	+
$3 - x$	+	+	−
f'	−	+	−
f	↘	↗	↘

From the table, we see that

- on the interval $[-3, 3]$, f is increasing;
- on the intervals $(-\infty, -3]$ and $[3, \infty)$, f is decreasing.

□

Remark

- In the last row of the above table, the first ↘ indicates that f is decreasing on $(-\infty, -3)$ etc. This information is obtained from Theorem 5.1.1.
- Since f is decreasing on $(-\infty, -3)$, by continuity, it is decreasing on $(-\infty, 3]$ etc.

Example Let $f : \mathbb{R} \longrightarrow \mathbb{R}$ be the function given by

$$f(x) = x^4 - 4x^3 + 5.$$

Find where the function f is increasing or decreasing.

Explanation This example is similar to the last one. The question is to find maximal intervals on which f is increasing or decreasing.

Solution Differentiating $f(x)$, we get

$$\begin{aligned} f'(x) &= \frac{\mathrm{d}}{\mathrm{d}x}(x^4 - 4x^3 + 5) \\ &= 4x^3 - 12x^2 \\ &= 4x^2(x - 3) \end{aligned}$$

	$(-\infty, 0)$	$(0, 3)$	$(3, \infty)$
4	+	+	+
x^2	+	+	+
$x - 3$	−	−	+
f'	−	−	+
f	↘	↘	↗

From the table, we see that

- on the interval $[3, \infty)$, f is increasing;
- on the interval $(-\infty, 3]$, f is decreasing.

□

Remark In the above example, to get the maximal interval on which f is decreasing, the following simple result is used.

Theorem 5.1.2 *Let f be a function that is defined on an open interval (a, c). If f is increasing (respectively decreasing) on (a, b) as well as on (b, c) and is continuous at b, then it is increasing (respectively decreasing) on (a, c).*

Proof We give the proof for the increasing case. Let $x_1, x_2 \in (a, c)$ and $x_1 < x_2$. We want to show that $f(x_1) < f(x_2)$. For this, we consider the following cases:

(1) If $x_1, x_2 \in (a, b)$, then $f(x_1) < f(x_2)$ since f is increasing on (a, b).

(2) If $x_1, x_2 \in (b, c)$, then $f(x_1) < f(x_2)$ since f is increasing on (b, c).

(3) If $x_1 \in (a, b)$ and $x_2 = b$, then $f(x_1) < f(x_2)$ since f is increasing on (a, b) and continuous at b.

(4) If $x_1 = b$ and $x_2 \in (b, c)$, then $f(x_1) < f(x_2)$ since f is increasing on (b, c) and continuous at b.

(5) If $x_1 \in (a, b)$ and $x_2 \in (b, c)$, then $f(x_1) < f(b) < f(x_2)$ by Cases (3) and (4).

□

5.1.2 Relative Extrema

In the last section, we use Theorem 5.1.1 to find where a function f is increasing or decreasing. For that, we solve inequalities $f'(x) > 0$ and $f'(x) < 0$. In each of the tables obtained in the last two examples, the intervals are obtained from $\mathbb{R}$ by deleting the zeros of f'. For example, in the last example, $\mathbb{R} \setminus \{0, 3\} = (-\infty, 0) \cup (0, 3) \cup (3, \infty)$. Zeros of f' are important because they play an important role in extremum problems.

Definition Let f be a function and let x_0 be a real number such that f is defined on an open interval containing x_0. If $f'(x_0) = 0$, then we say that x_0 is a *stationary number* of f.

Explanation

- If $x = t$ is time and $y = f(t)$ is the displacement (function) of a moving object, then $\frac{\mathrm{d}y}{\mathrm{d}t} = f'(t)$ is the velocity (function). Thus $f'(t_0) = 0$ means that the velocity at time t_0 is 0, that is, the object is stationary at that moment.
- In the definition, the condition "*f is defined on an open interval containing x_0*" can be omitted because the condition "$f'(x_0) = 0$" implicitly implies that f is defined on the left-side and right-side of x_0 as well as at x_0. However, we will continue to use this lengthy description to give the general setting.

Definition Let f be a function and let x_0 be a real number such that f is defined on an open interval containing x_0. If $f'(x_0)$ does not exist or $f'(x_0) = 0$, then we say that x_0 is a *critical number* of f.

Explanation

- Most functions considered in this course are differentiable (on open intervals that are subsets of their domains). For such functions, critical numbers and stationary numbers are the same.
- Instead of "*critical number*", many authors use the term "*critical point*". However, many students misunderstand this terminology and take $(x_0, f(x_0))$ as a critical point.

  Caution A critical point is a point on the real line (that is, a real number) rather than a point on the graph of f (an ordered pair).

Example Let $f : \mathbb{R} \longrightarrow \mathbb{R}$ be the function given by $f(x) = x^2 + 4x - 11$. Find the critical number(s) of f.

Explanation Because f is differentiable on $\mathbb{R}$, the question is to find the stationary numbers of f.

Solution Differentiating $f(x)$, we get
$$\begin{aligned} f'(x) &= \frac{\mathrm{d}}{\mathrm{d}x}(x^2 + 4x - 11) \\ &= 2x + 4. \end{aligned}$$

Solving $f'(x) = 0$, we get $x = -2$ which is the critical number of f. □

Definition Let f be a function and let x_0 be a real number such that f is defined on an open interval containing x_0. We say that

-] f has a *relative maximum* at $x = x_0$ if $f(x_0) \geq f(x)$ for all x sufficiently close to x_0;
- f has a *relative minimum* at $x = x_0$ if $f(x_0) \leq f(x)$ for all x sufficiently close to x_0.

Remark The condition "*$f(x_0) \geq f(x)$ for all x sufficiently close to x_0*" means that there exists an open interval (α, β) with $x_0 \in (\alpha, \beta)$ such that $f(x_0) \geq f(x)$ for all $x \in (\alpha, \beta)$. The interval (α, β) may be different from (a, b).

Terminology Suppose that f has a relative maximum (respectively relative minimum) at x_0. Then

- the number x_0 is called a *relative maximizer* (respectively *relative minimizer*) of f;
- the ordered pair $(x_0, f(x_0))$ is called a *relative maximum point* (respectively *relative minimum point*) of the graph of f;
- the number $f(x_0)$ is called a *relative maximum value* (respectively *relative minimum value*) of f.

Remark The terms *relative maximum value* and *relative minimum value* will not be used in this course because they are ambiguous; a value can be a relative maximum value as well as a relative minimum value.

If a function f has a relative maximum (respectively relative minimum) at x_0, then the graph of f has a peak (respectively a valley) at $(x_0, f(x_0))$, that is, the point $(x_0, f(x_0))$ is higher than (respectively lower than) its neighboring points. However, it may not be the highest point (respectively lowest point) on the whole graph. For this reason, we say that $(x_0, f(x_0))$ is a *local maximum point* (respectively *local minimum point*) of the graph.

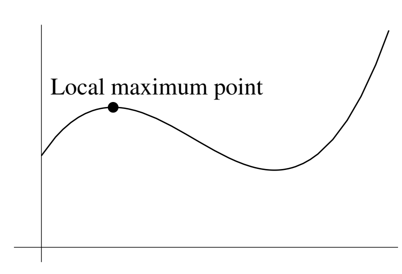


Figure 5.4

- The adjectives "*relative*" and "*local*" will be used interchangeably. For example, a local maximizer means a relative maximizer and a local minimum point means a relative minimum point etc.
- A local maximizer or a local minimizer will be called a *local extremizer* and a local maximum point or a local minimum point will be called a *local extremum point* etc.

Remark Consider the function $f : [0, 2] \longrightarrow \mathbb{R}$ defined by

$$f(x) = -2x^2 + 3x + 1.$$

The graph of f is shown in Figure 5.5. Although $f(0) \leq f(x)$ for all x sufficiently close to and greater 0, the number 0 is not considered as a local minimizer of f.

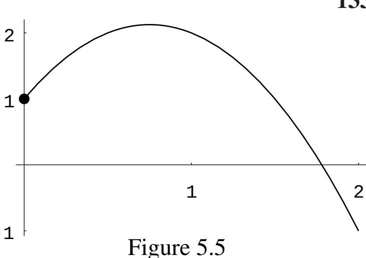


Figure 5.5

- In order to consider whether a number is a local extremizer of a function, the function has to be defined on an open interval containing the number.

The following theorem gives a necessary condition for local extremizers.

Theorem 5.1.3 *Let f be a function and let x_0 be a real number such that f is defined on an open interval containing x_0. Suppose that f has a local extremum at x_0. Then x_0 is a critical number of f, that is, $f'(x_0)$ does not exist or $f'(x_0) = 0$.*

Proof Suppose that $f'(x_0)$ exists (that is, f is differentiable at x_0). We want to show that $f'(x_0) = 0$.

Without loss of generality, we may assume that f has a local maximum at x_0 (otherwise, we may consider the function $-f$ instead). By definition, there exists an open interval (a, b) with $x_0 \in (a, b) \subseteq \operatorname{dom}(f)$ such that

$$f(x) \le f(x_0) \quad \text{for all } x \in (a, b). \tag{5.1.1}$$

Note that $f'(x_0) = \lim_{h\to 0} \frac{f(x_0 + h) - f(x_0)}{h}$; the limit exists and is a two-sided limit. Thus we have

$$f'(x_0) = \lim_{h\to 0-} \frac{f(x_0 + h) - f(x_0)}{h} = \lim_{h\to 0+} \frac{f(x_0 + h) - f(x_0)}{h}.$$

To show that $f'(x_0) = 0$, we show that one of the one-sided limits is at least zero and the other is at most zero.

- For $h < 0$ such that $a < x_0 + h$, by (5.1.1), we have $f(x_0 + h) \le f(x_0)$ which implies that
$$\frac{f(x_0 + h) - f(x_0)}{h} \ge 0.$$
Hence we have $\lim_{h\to 0-} \frac{f(x_0 + h) - f(x_0)}{h} \ge 0$.
- For $h > 0$ such that $x_0 + h < b$, by (5.1.1), we have $f(x_0 + h) \le f(x_0)$ which implies that
$$\frac{f(x_0 + h) - f(x_0)}{h} \le 0.$$
Hence we have $\lim_{h\to 0+} \frac{f(x_0 + h) - f(x_0)}{h} \le 0$.

Therefore, the required result follows. □

Remark For differentiable functions f, the result means that if f has a local extremum at x_0, then $f'(x_0) = 0$.

Example Let f be the function given by $f(x) = 2\sqrt{x} - x$. We want to apply Theorem 5.1.3 to look for all the possible local extremizers of f.

Note that the domain of f is $[0, \infty)$. Thus local extremizers of f must belong to $(0, \infty)$. For $x > 0$, by the Power Rule, we have $f'(x) = \frac{1}{\sqrt{x}} - 1$. Thus f is differentiable on $(0, \infty)$. To look for local extremizers of f, by Theorem 5.1.3, we only need to find positive real numbers x_0 such that $f'(x_0) = 0$. Solving $\frac{1}{\sqrt{x}} - 1 = 0$, we get $x = 1$, which is the only possible candidate for local extremizer of f.

The next example shows that the converse Theorem 5.1.3 is not true: if $f'(x_0) = 0$, f may not have a local extremum at x_0.

Example Let $f : \mathbb{R} \longrightarrow \mathbb{R}$ be the function given by

$$f(x) = x^3 - 3x^2 + 3x.$$

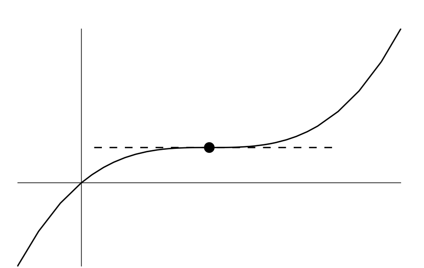
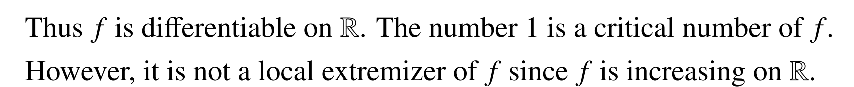


Figure 5.6

Then we have
$$\begin{aligned} f'(x) &= \frac{\mathrm{d}}{\mathrm{d}x}(x^3 - 3x^2 + 3x) \\ &= 3x^2 - 6x + 3 \\ &= 3(x-1)^2. \end{aligned}$$



Thus f is differentiable on $\mathbb{R}$. The number 1 is a critical number of f. However, it is not a local extremizer of f since f is increasing on $\mathbb{R}$.

Suppose x_0 is a critical number of a function f. At x_0, the function f may have a local maximum, a local minimum or neither. The next result describes a simple way to determine which case it is using the first derivative of f.

First Derivative Test Let f be a function that is differentiable on an open interval (a, b) and let $x_0 \in (a, b)$. Suppose that x_0 is a critical number of f.

(1) If $f'(x)$ changes from positive to negative as x increases through x_0, then x_0 is a local maximizer of f.

(2) If $f'(x)$ changes from negative to positive as x increases through x_0, then x_0 is a local minimizer of f.

(3) If $f'(x)$ does not change sign as x increases through x_0, then x_0 is neither a local maximizer nor local minimizer of f.

Explanation The assumption on x_0 is that $f'(x_0) = 0$.

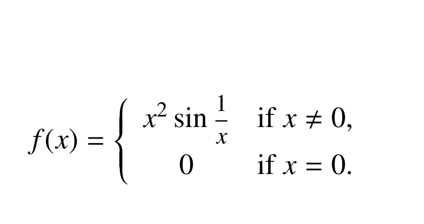
- The condition "$f'(x)$ *changes from positive to negative as x increases through* x_0" means that $f'(x) > 0$ for x sufficiently close to and less than x_0 and $f'(x) < 0$ for x sufficiently close to and greater than x_0.
- The condition "$f'(x)$ *does not change sign as x increases through* x_0" means that $f'(x)$ is either always positive or always positive for x sufficiently close and different from x_0.

Proof We give the proof for (1) and (3). The proof for (2) is similar to that for (1).

(1) If $f'(x)$ changes from positive to negative as x increases through x_0, then there is an open interval in the form (a, x_0) such that $f'(x) > 0$ for all $x \in (a, x_0)$ and there is an open interval in the form (x_0, b) such that $f'(x) < 0$ for all $x \in (x_0, b)$; hence by Theorem 5.1.1 and continuity, f is increasing on $(a, x_0]$ and decreasing on $[x_0, b)$. Therefore, f has a local maximum at x_0.

(3) If $f'(x)$ does not change sign as x increases through x_0, then there are open intervals in the form (a, x_0) and (x_0, b) such that $f'(x) > 0$ for all $x \in (a, x_0) \cup (x_0, b)$ or $f'(x) < 0$ for all $x \in (a, x_0) \cup (x_0, b)$. In the first case, by Theorem 5.1.1 f is increasing on (a, x_0) as well as on (x_0, b) and hence by continuity, it is increasing on (a, b). In the second case, f is decreasing on (a, b). Therefore, in any case, f does not have a local extremum at x_0. □

Remark For "*nice*" functions (for example, polynomials), the above result includes all possibilities. But we can construct weird functions f such that x_0 is a critical number of f and that f' changes sign infinitely often on the left and right of x_0. Figure 5.7 shows the graph of the function f given below; the number 0 is a critical number of f.



$$f(x) = \begin{cases} x^2 \sin \frac{1}{x} & \text{if } x \neq 0, \\ 0 & \text{if } x = 0. \end{cases}$$

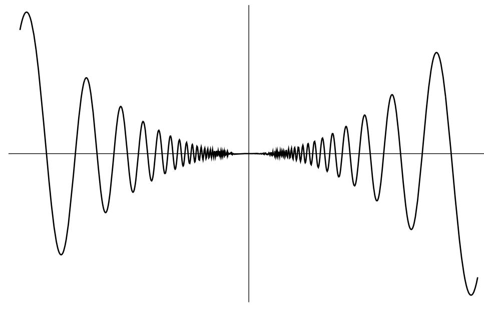


Figure 5.7

Example Let $f : \mathbb{R} \longrightarrow \mathbb{R}$ be the function given by

$$f(x) = 27x - x^3.$$

Find and determine the nature of the critical number(s) of f.

Explanation The question is to find all the critical numbers of f and for each critical number, determine whether it is a local maximizer, a local minimizer or not a local extremizer.

Solution Differentiating $f(x)$, we get

$$\begin{aligned} f'(x) &= \frac{\mathrm{d}}{\mathrm{d}x}(27x - x^3) \\ &= 27 - 3x^2 \\ &= 3(3 + x)(3 - x). \end{aligned}$$

Solving $f'(x) = 0$, we get the critical numbers of f: $x_1 = -3$ and $x_2 = 3$.

- When x is sufficiently close to and less than -3, $f'(x)$ is negative; when x is sufficiently close to and greater than -3, $f'(x)$ is positive. Hence, by the First Derivative Test, $x_1 = -3$ is a local minimizer of f.
- When x is sufficiently close to and less than 3, $f'(x)$ is positive; when x is sufficiently close to and greater than 3, $f'(x)$ is negative. Hence, by the First Derivative Test, $x_2 = 3$ is a local maximizer of f.

□

Remark The function in the above example is considered in the last subsection in which a table is obtained. It is clear from the table that f has a local minimum at -3 and a local maximum at 3. In the next example, we will use the table method to determine nature of critical numbers.

	$(-\infty, -3)$	$(-3, 3)$	$(3, \infty)$
3	+	+	+
$3 + x$	−	+	+
$3 - x$	+	+	−
f'	−	+	−
f	↘	↗	↘

Example Let $f : \mathbb{R} \longrightarrow \mathbb{R}$ be the function given by

$$f(x) = x^4 - 4x^3 + 5.$$

Find and determine the nature of the critical number(s) of the f.

Explanation The function is considered in an example in the last subsection. Below we just copy the main steps from the solution there.

Solution

$$\begin{aligned} f'(x) &= 4x^3 - 12x^2 \\ &= 4x^2(x - 3) \end{aligned}$$

	$(-\infty, 0)$	$(0, 3)$	$(3, \infty)$
f'	−	−	+
f	↘	↘	↗

Solving $f'(x) = 0$, we get the critical numbers of f: $x_1 = 0$ and $x_2 = 3$.

From the table, we see that

- the critical number $x_1 = 0$ is not a local extremizer of f;
- the critical number $x_2 = 3$ is a local minimizer of f.

□

5.1.3 Convexity

In studying curves, we are also interested in finding out how the curves bend. Both curves shown in Figures 5.8(a) and (b) go up (as x goes from left to right). However the way how they bend are quite different.

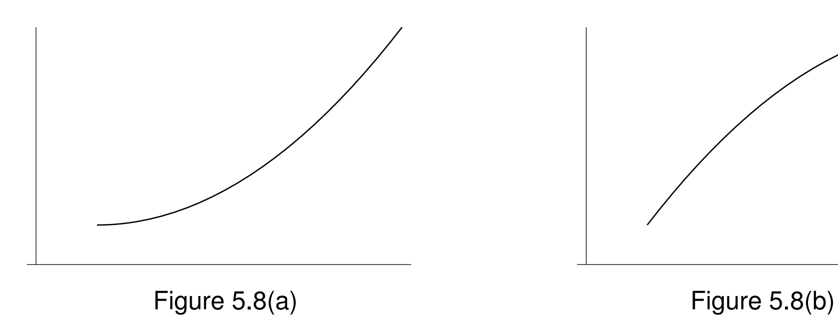


Figure 5.8(a) Figure 5.8(b)

- In Figure 5.8(a), the curve goes up faster and faster, that is, the slope becomes more and more positive as we move from left to right (the slope is increasing). We say that the curve is *bending up*.
- In Figure 5.8(b), although the curve goes up, the slope becomes less and less positive (the slope is decreasing). We say that the curve is *bending down*.

Similarly we can consider curves that go down.

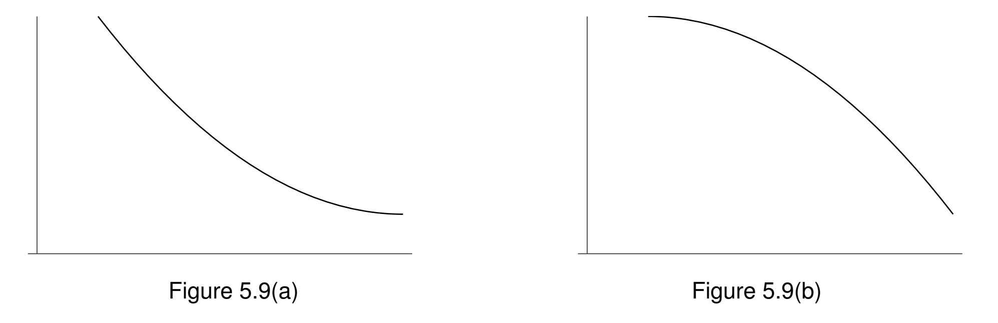


Figure 5.9(a) Figure 5.9(b)

- The curve in Figure 5.9(a) goes down. However, the slope becomes less and less negative. This means that the slope is increasing and we say that the curve is *bending up*.
- The curve in Figure 5.9(b) also goes down. Moreover, the slope becomes more and more negative. This means that the slope is decreasing and we say that the curve is *bending down*.

In summary, curves having shape shown in Figure 5.10(a) [*or part of it*] is said to be *bending up* and those having shape shown in Figure 5.10(b) [*or part of it*] is said to be *bending down*.

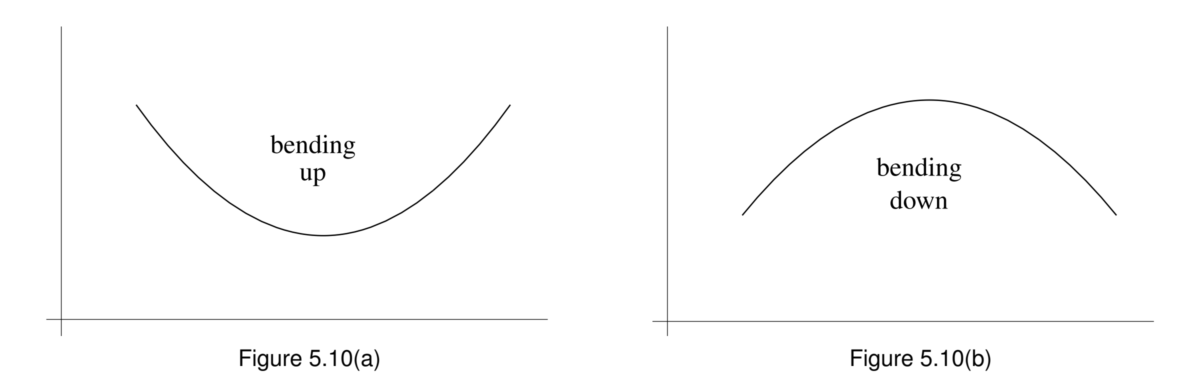


Figure 5.10(a) Figure 5.10(b)

Alternatively, if the curve is the graph of $y = f(x)$ where f is a differentiable function, the graph is bending up (respectively bending down) means that the graph is always above (respectively always below) the tangent lines.

Remark In many books, instead of bending up and bending down, the terms *concave up* and *concave down* respectively are used.

Bending up and bending down are properties of curves. Below are properties of functions corresponding to these geometric properties.

Definition Let f be a function that is defined and differentiable on an open interval (a, b). We say that

- f is *strictly convex* on (a, b) if f' is increasing on (a, b);
- f is *strictly concave* on (a, b) if f' is decreasing on (a, b).

Since f' is the slope function, f is strictly convex on (a, b) means that the slope is increasing and so in the interval (a, b), the graph of f is bending up. Similarly, f is strictly concave means that in (a, b), its graph is bending down.

Terminology For simplicity, instead of saying "*strictly convex*", we will say "*convex*" etc.

The next theorem describes a simple way to find where a function is convex or concave. The method is to consider the sign of f''.

Theorem 5.1.4 *Let f be a function that is defined and is twice differentiable on an open interval (a, b).*

(1) *If $f''(x) > 0$ for all $x \in (a, b)$, then f is convex on (a, b).*

(2) *If $f''(x) < 0$ for all $x \in (a, b)$, then f is concave on (a, b).*

Proof We give the proof for (1). The proof of (2) is similar to that for (1).

Note that f'' is the derivative of f'. If $f''(x) > 0$ for all $x \in I$, that is, $(f')'(x) > 0$ for all $x \in (a, b)$, then by Theorem 5.1.1, f' is increasing on (a, b), that is, f is convex on (a, b). □

Example Let $f : \mathbb{R} \longrightarrow \mathbb{R}$ be the function given by

$$f(x) = 27x - x^3.$$

Find the interval(s) on which f is convex or concave.

Explanation

- The question is to find maximal open interval(s), if any, on which f is convex or concave.
- The given function f is a "nice" function (a polynomial function). It can be differentiated any number of times: $f^{(n)}(x)$ exists for all positive integers n and for all real numbers x. In particular, f is twice differentiable on $\mathbb{R}$. To apply Theorem 5.1.4, we have to solve inequalities $f''(x) > 0$ and $f''(x) < 0$. This is done by setting up a table.

Solution Differentiating $f(x)$, we get
$$f'(x) = \frac{\mathrm{d}}{\mathrm{d}x}(27x - x^3) = 27 - 3x^2.$$

Differentiating $f'(x)$, we get
$$f''(x) = \frac{\mathrm{d}}{\mathrm{d}x}(27 - 3x^2) = -6x.$$

	$(-\infty, 0)$	$(0, \infty)$
-6	$-$	$-$
x	$-$	$+$
f''	$+$	$-$

- On the interval $(-\infty, 0)$, f is convex.
- On the interval $(0, \infty)$, f is concave.

□

Remark

- When we consider "*a function is convex/concave on an interval*", unlike increasing/decreasing, we do not include the endpoint(s) of the interval. This is because the concept is defined for open intervals only. Note that for a function f whose domain is a closed and bounded interval $[a, b]$, $f'(x)$ is undefined when $x = a$ or b.
- There is a more general definition for convex/concave functions. The definition does not involve f' and it can be applied to closed intervals also.

Example Let $f : \mathbb{R} \longrightarrow \mathbb{R}$ be the function given by

$$f(x) = x^4 - 4x^3 + 5.$$

Find where the graph of f is bending up or bending down.

Explanation This question is similar to the last one. The graph of f is bending up (or down) means that f is convex (or concave). So we have to find intervals on which f'' is positive (or negative).

Solution Differentiating $f(x)$, we get
$$f'(x) = \frac{\mathrm{d}}{\mathrm{d}x}(x^4 - 4x^3 + 5) = 4x^3 - 12x^2.$$

Differentiating $f'(x)$, we get
$$f''(x) = \frac{\mathrm{d}}{\mathrm{d}x}(4x^3 - 12x^2) = 12x^2 - 24x = 12x\,(x - 2).$$

	$(-\infty, 0)$	$(0, 2)$	$(2, \infty)$
$12x$	$-$	$+$	$+$
$x - 2$	$-$	$-$	$+$
f''	$+$	$-$	$+$

- On the intervals $(-\infty, 0)$ and $(2, \infty)$, the graph of f is bending up.

- On the interval $(0, 2)$, the graph of f is bending down.

□

Definition Let f be a function and let x_0 be a real number such that f is continuous at x_0 and differentiable on both sides of x_0. If f is convex on one side of x_0 and concave on the other side, then we say that x_0 is an *inflection number* of f.

Explanation

- The condition "*f is differentiable on both sides of x_0*" means that there is an open interval in the form (a, x_0) and an open interval in the form (x_0, b) such that $f'(x)$ exists for all $x \in (a, x_0) \cup (x_0, b)$.
- The condition "*f is convex on one side of x_0 and concave on the other side*" means that there is an open interval in the form (α, x_0) and an open interval in the form (x_0, β) on which f is convex on one of them and concave on the other, that is, there is a change of convexity at x_0.

Suppose that x_0 is an inflection number of a function f. By definition, on one side of the point $(x_0, f(x_0))$, the graph of f is bending up and on the other side, the graph is bending down. That is, there is a change of bending at the point $(x_0, f(x_0))$.

Terminology Suppose that x_0 is an inflection number of a function f. Then the point $(x_0, f(x_0))$ is called an *inflection point* of the graph of f.

In the following example, the function f is discussed in a previous example. Below we just copy part of the table obtained in the solution there.

Example Consider the function $f : \mathbb{R} \longrightarrow \mathbb{R}$ given by

$$f(x) = x^4 - 4x^3 + 5.$$

From the table

	$(-\infty, 0)$	$(0, 2)$	$(2, \infty)$
f''	$+$	$-$	$+$

we see that the inflection numbers of f are 0 and 2.

Remark We can also say that the inflection points of the graph are $(0, 5)$ and $(2, -11)$.

The next result gives a necessary condition for inflection number.

Theorem 5.1.5 *Let f be a function and let x_0 be a real number such that f is differentiable on an open interval containing x_0 and that $f''(x_0)$ exists. Suppose that x_0 is an inflection number of f. Then we have $f''(x_0) = 0$.*

Proof By symmetry, we may assume that f is convex on the left-side of x_0 and concave on the right-side of x_0, that is, there exist real numbers a and b with $a < x_0 < b$ such that f' is increasing on (a, x_0) and decreasing on (x_0, b), which by continuity of f' at x_0, implies that

$$f'(x) < f'(x_0) \quad \text{for all } x \in (a, x_0) \cup (x_0, b).$$

Thus, the function f' has a local maximum at x_0. Hence by Theorem 5.1.3 (and using the assumption that the derivative of f' at x_0 exists), the derivative of f' at x_0 is 0, that is, $f''(x_0) = 0$. □

Remark The converse of Theorem 5.1.5 is not true: if $f''(x_0) = 0$, x_0 may not be an inflection number of f.

Example Let $f : \mathbb{R} \longrightarrow \mathbb{R}$ be the function given by

$$f(x) = x^4.$$

Then we have

$$\begin{aligned} f'(x) &= 4x^3 \\ f''(x) &= 12x^2 \end{aligned}$$

	$(-\infty, 0)$	$(0, \infty)$
f''	+	+

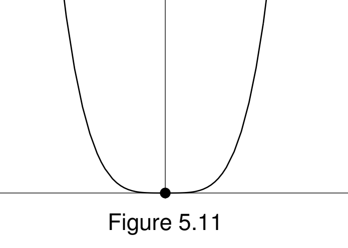


Figure 5.11

Although $f''(0) = 0$, the number 0 is not an inflection number of f. This is because f is convex on $(-\infty, 0)$ as well as on $(0, \infty)$.

Remark The function f is convex on $(-\infty, \infty)$.

Terminology

- If $f'(x_0) = 0$, we say that x_0 is a stationary number of f. However, if $f''(x_0) = 0$, we do not have a specific name for x_0.
- For local extremizers, there are two types: local maximizers and local minimizers. Correspondingly, there are also two types of inflection numbers. However, we do not have specific names to distinguish the two types.

Example Let $f : \mathbb{R} \longrightarrow \mathbb{R}$ be the function given by

$$f(x) = x^4 - 6x^2 + 5x - 6.$$

Find the inflection point(s) of the graph of f.

Explanation To find the inflection points of the graph, first we find the inflection numbers of the function. For that, we solve the equation $f''(x) = 0$. By Theorem 5.1.5, solutions to this equation include all the possible candidates for inflection numbers. However, for each of these candidates, we have to check whether the convexity of f are different on the left-side and right-side of it.

Solution Differentiating $f(x)$, we get

$$\begin{aligned} f'(x) &= \frac{\mathrm{d}}{\mathrm{d}x}(x^4 - 6x^2 + 5x - 6) \\ &= 4x^3 - 12x + 5. \end{aligned}$$

Differentiating $f'(x)$, we get

$$\begin{aligned} f''(x) &= \frac{\mathrm{d}}{\mathrm{d}x}(4x^3 - 12x + 5) \\ &= 12x^2 - 12 \\ &= 12(x + 1)(x - 1). \end{aligned}$$

Solving $f''(x) = 0$, we get two solutions: $x_1 = 1$ and $x_2 = -1$.

	$(-\infty,-1)$	$(-1,1)$	$(1,\infty)$
12	+	+	+
$x+1$	−	+	+
$x-1$	−	−	+
f''	+	−	+

From the table, we see that f is convex on $(-\infty,-1)$, concave on $(-1,1)$ and convex on $(1,\infty)$. Hence $x_1 = 1$ and $x_2 = -1$ are the inflection numbers of f.

The inflection points of the graph are $(1, f(1)) = (1, -6)$ and $(-1, f(-1)) = (-1, -16)$. □

To determine the nature of critical numbers, we can use the First Derivative Test discussed in the last subsection. Below, we discuss an alternative way using second derivatives.

Second Derivative Test Let f be a function and let x_0 be a real number such that f is differentiable on an open interval containing x_0. Suppose that x_0 is a critical number of f, that is, $f'(x_0) = 0$.

(1) If $f''(x_0) < 0$, then x_0 is a local maximizer of f (in fact, we have $f(x_0) > f(x)$ for all x sufficiently close to and different from x_0).

(2) If $f''(x_0) > 0$, then x_0 is a local minimizer of f (in fact, we have $f(x_0) < f(x)$ for all x sufficiently close to and different from x_0).

Explanation Below we give a proof for (1). To prove (2), we can use the method for (1). Alternatively, we can apply (1) to the function $-f$ because $(-f)''(x_0) < 0$ in this case.

Proof It suffices to prove (1). Suppose that $f''(x_0) < 0$. We want to show that f is increasing on the left-side of x_0 and decreasing on the right-side.

By definition, together with the condition $f'(x_0) = 0$, we have

$$0 > f''(x_0) = \lim_{h\to 0} \frac{f'(x_0+h)}{h},$$

which implies that $f'(x_0 + h) > 0$ if h is sufficiently close to and less than 0, that is,

$$f'(x) > 0 \quad \text{if } x = x_0 + h \text{ is sufficiently close to and less than } x_0.$$

Hence, by Theorem 5.1.1, f is increasing on the left-side of x_0. Similarly, f is decreasing on the right-side of x_0. Therefore, by the continuity of f at x_0, we see that $f(x_0) > f(x)$ for all x sufficiently close to and different from x_0. □

Remark

- To determine the nature of a critical number using the Second Derivative Test, we consider the sign of f'' *at the critical number*. If we apply the First Derivative Test, we consider the sign of f' *on the left-side and the right-side of the critical number*.
- If $f''(x_0) = 0$, we can't apply the Second Derivative Test. At x_0, the function f may have a local maximum, a local minimum or neither. See the last example in this subsection.

- Some students have the following conjecture:

  Suppose $f'(x_0) = 0$ *and* x_0 *is not a local extremizer of* f*, then* x_0 *is an inflection number of* f.

  For "nice" functions (for example, polynomial functions), the conjecture is correct. However, we can construct weird functions with "weird" critical point (see Figure 5.7).

  However, when we consider nature of a critical number, there is no need to discuss whether it is an inflection number because critical numbers are related to first derivatives whereas inflection numbers are related to second derivatives.

Below, we redo a previous example using the Second Derivative Test.

Example Let $f : \mathbb{R} \longrightarrow \mathbb{R}$ be the function given by

$$f(x) = 27x - x^3.$$

Find and determine the nature of the critical number(s) of f.

Solution Differentiating $f(x)$, we get

$$\begin{aligned} f'(x) &= \frac{\mathrm{d}}{\mathrm{d}x}(27x - x^3) \\ &= 27 - 3x^2 \\ &= 3(3 + x)(3 - x). \end{aligned}$$

Solving $f'(x) = 0$, we get the critical numbers of f: $x_1 = -3$ and $x_2 = 3$.

Differentiating $f'(x)$, we get

$$\begin{aligned} f''(x) &= \frac{\mathrm{d}}{\mathrm{d}x}(27 - 3x^2) \\ &= -6x \end{aligned}$$

- At $x_1 = -3$, we have $f''(-3) = 18 > 0$; therefore, x_1 is a local minimizer of f.
- At $x_2 = 3$, we have $f''(3) = -18 < 0$; therefore, x_2 is a local maximizer of f.

□

Example Let f, g and h be functions from $\mathbb{R}$ to $\mathbb{R}$ given by

$$f(x) = x^4, \qquad g(x) = -x^4, \qquad h(x) = x^3.$$

It is clear that $x_1 = 0$ is a critical number of f, g and h. Moreover, we have $f''(0) = g''(0) = h''(0)$. However,

- at $x_1 = 0$, f has a local minimum;
- at $x_1 = 0$, g has a local maximum;
- at $x_1 = 0$, h does not have a local extremum.

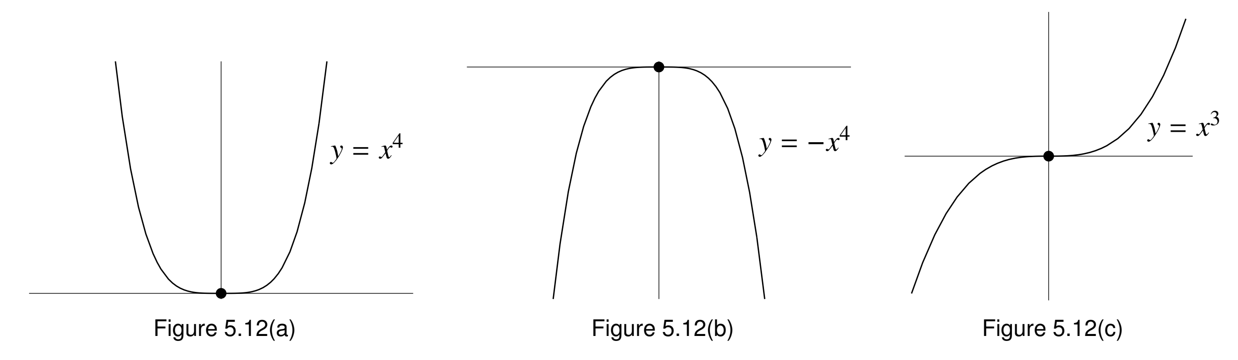


Figure 5.12(a) Figure 5.12(b) Figure 5.12(c)

5.1.4 Curve Sketching

Given a function f that is twice differentiable on an open interval (a, b), to sketch the graph of $y = f(x)$ for $a < x < b$, we can use the first derivative of f to find where the graph goes up or down and use the second derivative of f to find where the graph bends up or down. Hence we can locate the local extremum points and inflection points of the graph. Intercepts give additional information for the graph. If f is a rational function, limits at infinity ($\pm\infty$) and vertical asymptotes (infinite limits) are also useful.

The following table gives the shape of the graph of f corresponding to the four cases determined by the signs of f' and f''. For example, first row first column corresponds to that both f' and f'' are positive: the figure indicates that the graph goes up and bends up.

	$f' > 0$	$f' < 0$
$f'' > 0$		
$f'' < 0$		

Example Sketch the graph of $y = 27x - x^3$ for $x \in [-5.5, 5.5]$.

Explanation In the question, we are ask to draw the graph of f where $f(x) = 27x - x^3$ for $-5.5 \leq x \leq 5.5$. In the graph, we should locate the endpoints $(-5.5, f(-5.5))$ and $(5.5, f(5.5))$. In two previous examples, we obtain the following:

	$(-\infty, -3)$	$(-3, 3)$	$(3, \infty)$
f'	$-$	$+$	$-$

	$(-\infty, 0)$	$(0, \infty)$
f''	$+$	$-$

The three numbers -3, 3 (zeros of f') and 0 (zero of f'') divide $\mathbb{R}$ into four intervals: $(-\infty, -3)$, $(-3, 0)$, $(0, 3)$ and $(3, \infty)$. On each of these intervals, we can use the above tables to consider the signs of f' and f''.

Solution

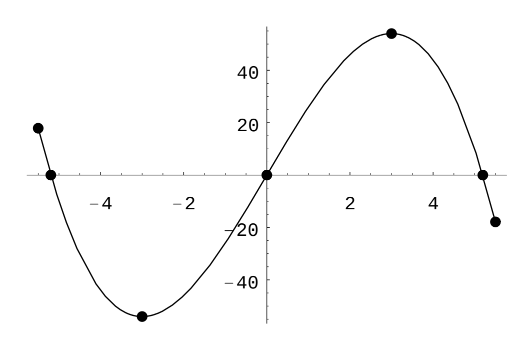
	$(-\infty, -3)$	$(-3, 0)$	$(0, 3)$	$(3, \infty)$
f'	$-$	$+$	$+$	$-$
f''	$+$	$+$	$-$	$-$
f				

On the graph, we have

- Local minimum point $\quad (-3, f(-3)) = (-3, -54)$
- Inflection point $\quad (0, f(0)) = (0, 0)$

- Local maximum point $(3, f(3)) = (3, 54)$
- Intercepts $(0, 0)$, $(3\sqrt{3}, 0)$ and $(-3\sqrt{3}, 0)$
- Endpoints $(-5.5, f(-5.5)) = (-5.5, 17.875)$
  and $(5.5, f(5.5)) = (5.5, -17.875)$

The required graph is shown in the following figure:



□

Remark Since f is an odd function, the graph is symmetric about the origin.

Example Sketch the graph of $y = x^4 - 4x^3 + 5$ for $-1.5 \leq x \leq 4.2$.

Explanation In two previous examples, we obtain the following:

	$(-\infty, 0)$	$(0, 3)$	$(3, \infty)$
f'	$-$	$-$	$+$

	$(-\infty, 0)$	$(0, 2)$	$(2, \infty)$
f''	$+$	$-$	$+$

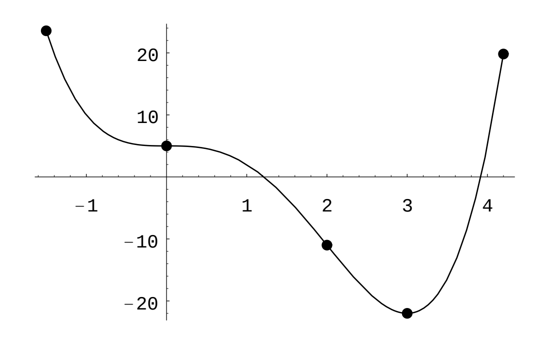
Solution

	$(-\infty, 0)$	$(0, 2)$	$(2, 3)$	$(3, \infty)$
f'	$-$	$-$	$-$	$+$
f''	$+$	$-$	$+$	$+$
f	↘ (concave up)	↘ (concave down)	↘ (concave up)	↗ (concave up)

On the graph, we have

- Inflection points $(0, f(0)) = (0, 5)$ and $(2, f(2)) = (2, -11)$
- Local minimum point $(3, f(3)) = (3, -22)$
- Endpoints $(-1.5, f(-1.5)) \approx (-1.5, 23.6)$
  and $(4.2, f(4.2)) \approx (4.2, 19.8)$

The required graph is shown in the following figure:



□

Remark There are two x-intercepts. Approximate values of their x-coordinates are 1.2 and 3.9 which can be estimated using the Intermediate Value Theorem.

Example Sketch the graph of $y = x^3 + 3x^2 - 45x$ for $-9 \le x \le 6$.

Solution Differentiating $f(x)$, we get

$$\begin{aligned} f'(x) &= 3x^2 + 6x - 45 \\ &= 3(x-3)(x+5) \end{aligned}$$

	$(-\infty, -5)$	$(-5, 3)$	$(3, \infty)$
3	+	+	+
$x - 3$	−	−	+
$x + 5$	−	+	+
f'	+	−	+

Differentiating $f'(x)$, we get $f''(x) = 6x + 6$

	$(-\infty, -1)$	$(-1, \infty)$
f''	−	+

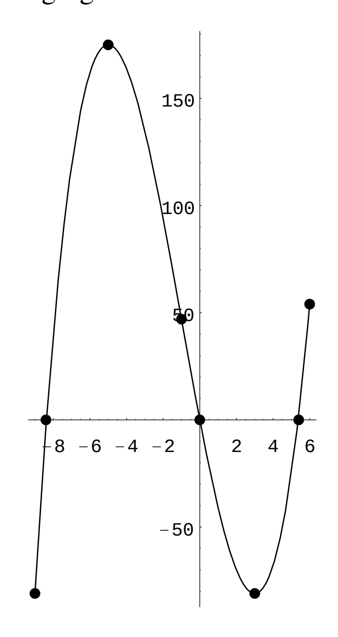
Combining the two tables, we get

	$(-\infty, -5)$	$(-5, -1)$	$(-1, 3)$	$(3, \infty)$
f'	+	−	−	+
f''	−	−	+	+
f	╭	╮	╰	╯

On the graph, we have

- Local maximum point $(-5, f(-5)) = (-5, 175)$
- Inflection point $(-1, f(-1)) = (-1, 47)$
- Local minimum point $(3, f(3)) = (3, -81)$
- Intercepts $(0, 0)$, $(\frac{-3 + 3\sqrt{21}}{2}, 0)$ and $(\frac{-3 - 3\sqrt{21}}{2}, 0)$
- Endpoints $(-9, f(-9)) = (-9, -81)$ and $(6, f(6)) = (6, 54)$

The required graph is shown in the following figure:



□

Exercise 5.1

1. For each of the following functions f, find the interval(s) on which it is increasing.
   (a) $f(x) = 2x^2 - 5x + 6$ (b) $f(x) = 1 + 3x - x^3$
   (c) $f(x) = x^3 + 6x^2 - 63x$ (d) $f(x) = 2x^3 + 9x^2 - 6x + 7$
   (e) $f(x) = 3x^4 + 4x^3 - 24x^2 - 48x$ (f) $f(x) = x + \frac{4}{x}$
2. For each of the following functions f, find and determine the nature of its critical number(s).
   (a) $f(x) = -x^2 + 7x - 13$ (b) $f(x) = x^4 - 2x^3$
   (c) $f(x) = x^5 - 15x^3$ (d) $f(x) = \frac{x^2 + x + 1}{x + 1}$
3. For each of the following functions f, find the interval(s) on which it is convex.
   (a) $f(x) = \sqrt{x}$ (b) $f(x) = x^3 - 6x^2 + 9x$
   (c) $f(x) = 3x^5 - 9x^4 + 8x^3$ (d) $f(x) = x + \frac{2}{x}$
4. For each of the following functions f, find its inflection number(s).
   (a) $f(x) = 2x^3 + 9x^2 - 108x + 35$ (b) $f(x) = 1 - \frac{1}{x} + \frac{1}{x^2}$
5. For each of the following equations, sketch its graph (*you have to choose a suitable interval*).
   (a) $y = x^3 - 6x^2$ (b) $y = 8x^3 - 2x^4$
   (c) $y = (x^2 - 3)^2$ (d) $y = x^3 + x + 1$

5.2 Applied Extremum Problems

In this section, we will consider application of differentiation to applied extremum problems. In such problems, we are interested in *absolute* (or *global*) extrema rather than *relative* extrema.

5.2.1 Absolute Extrema

Definition Let f be a function and let x_0 be a real number belonging to the domain of f.

- If $f(x_0) \geq f(x)$ for all $x \in \operatorname{dom}(f)$, then we say that f attains its (*absolute* or *global*) *maximum* at x_0 and that the number $f(x_0)$ is the (*absolute* or *global*) *maximum* (*value*) of f.
- If $f(x_0) \leq f(x)$ for all $x \in \operatorname{dom}(f)$, then we say that f attains its (*absolute* or *global*) *minimum* at x_0 and that the number $f(x_0)$ is the (*absolute* or *global*) *minimum* (*value*) of f.

Remark Maximum and minimum values are unique (if exist).

Example Let $f : \mathbb{R} \longrightarrow \mathbb{R}$ be the function given by $f(x) = x^2 + 1$. Then

- f attains its (absolute) minimum at 0 and the minimum of f is 1;
- f does not attain its (absolute) maximum, that is, there does not exist any $x_0 \in \mathbb{R}$ such that $f(x_0) \geq f(x)$ for all $x \in \mathbb{R}$.

Terminology Maximum and minimum (values) of a function f are called (*absolute* or *global*) *extrema* of f.

Recall (Extreme Value Theorem) *Let $f : [a, b] \longrightarrow \mathbb{R}$ be a continuous function. Then f attains its (absolute) maximum and minimum. That is, there exist $x_1, x_2 \in [a, b]$ such that*

$$f(x_1) \leq f(x) \leq f(x_2) \qquad \text{for all } x \in [a, b].$$

Note Extrema may occur at the endpoints a, b or at points in (a, b).

Figure 5.13 shows the graph of a function f with domain $[a, b]$. Note that f attains its absolute minimum at x_2 which belongs to the open interval (a, b) and attains its absolute maximum at b which is an endpoint. Also note that f has a relative maximum at x_1 but it does not attains its absolute maximum there.

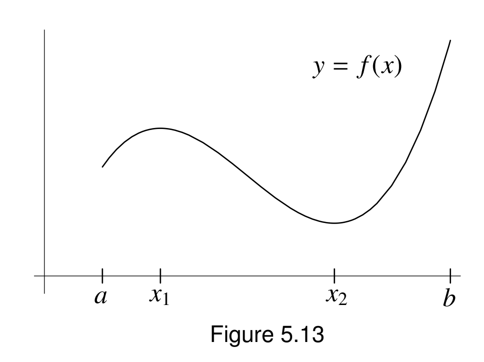


Figure 5.13

Let $f : [a, b] \longrightarrow \mathbb{R}$ be a function that is differentiable on (a, b). Suppose that f attains its maximum or minimum at x_0 where $a < x_0 < b$. Then by Theorem 5.1.3, x_0 must be a critical number of f, that is, $f'(x_0) = 0$. Thus we have the following procedures to find the absolute extrema of f.

Steps to find absolute extrema

(1) Find the critical number(s) of f in (a, b).

(2) Find the values of f at the endpoints a and b and that at the critical number(s) found in (1).

(3) The maximum and minimum values of f are, respectively, the greatest and smallest of the values found in Step 2.

FAQ Do we need to check the nature (relative maximum or minimum) of the critical numbers?

Answer If you want to find absolute extrema, there is no need to check the nature of the critical numbers. Even if you know that f has a local maximum (say) at a certain critical number x_0, you still have to compare values. However if you know that f is increasing on $[a, x_0]$ and decreasing on $[x_0, b]$, then you can tell that f attains its absolute maximum at x_0, that is, $f(x_0)$ is the absolute maximum; and to get the absolute minimum, you can compare the values $f(a)$ and $f(b)$. □

Example Find the absolute extremum values of the function f given by

$$f(x) = 2x^3 - 18x^2 + 30x$$

on the closed interval $[0, 3]$.

Explanation In this question, the domain of f is taken to be $[0, 3]$. Since f is continuous on $[0, 3]$, it follows from the Extreme Value Theorem that f attains its absolute extrema. Note that f is differentiable on $(0, 3)$. Thus we can apply the above steps to find the absolute extremum values.

Solution Differentiating $f(x)$, we get

$$\begin{aligned} f'(x) &= \frac{\mathrm{d}}{\mathrm{d}x}(2x^3 - 18x^2 + 30x) \\ &= 6x^2 - 36x + 30 \qquad (0 < x < 3) \end{aligned}$$

Solving $f'(x) = 0$, that is,

$$\begin{aligned} 6x^2 - 36x + 30 &= 0 \quad (0 < x < 3) \\ 6(x-1)(x-5) &= 0 \quad (0 < x < 3) \end{aligned}$$

we get the critical number of f in $(0, 3)$: $x_1 = 1$.

Comparing the values of f at the critical number and that at the endpoints:

x	0	1	3
$f(x)$	0	14	−18

we see that the maximum value of f is 14 and the minimum value of f is −18. □

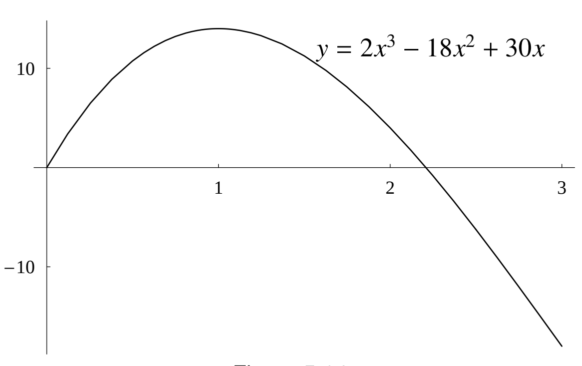


Figure 5.14

5.2.2 Applied Maxima and Minima

Example An article in a sociology journal stated that if a particular health-care program for the elderly were initiated, then t years after its start, n thousand elderly people would receive direct benefits, where

$$n(t) = \frac{t^3}{3} - 6t^2 + 32t, \quad 0 \le t \le 10.$$

After how many years does the number of people receiving benefits attain maximum?

Solution Differentiating $n(t)$, we get

$$\begin{aligned} n'(t) &= \frac{d}{dt}\Big(\frac{t^3}{3} - 6t^2 + 32t\Big) \\ &= t^2 - 12t + 32 \qquad (0 < t < 10) \end{aligned}$$

Solving $n'(t) = 0$, that is,

$$\begin{aligned} t^2 - 12t + 32 &= 0 \quad (0 < t < 10) \\ (t-8)(t-4) &= 0 \quad (0 < t < 10), \end{aligned}$$

we get the critical numbers of n in $(0, 10)$: $t_1 = 4$ and $t_2 = 8$.

Comparing the values of n at the critical numbers and that at the endpoints:

x	0	4	8	10
$n(x)$	0	$\frac{160}{3}$	$\frac{128}{3}$	$\frac{160}{3}$

we see that n attains its maximum at $t_1 = 4$ and also at $t_2 = 10$.

The number of people receiving benefits attains maximum after 4 years as well as after 10 years. □

Remark Although the maximum (if exist) of a function is unique, the above example shows that the values of x at which a function attains its maximum may not be unique. The following figure show the graph of the function n. Note that there are two highest points.

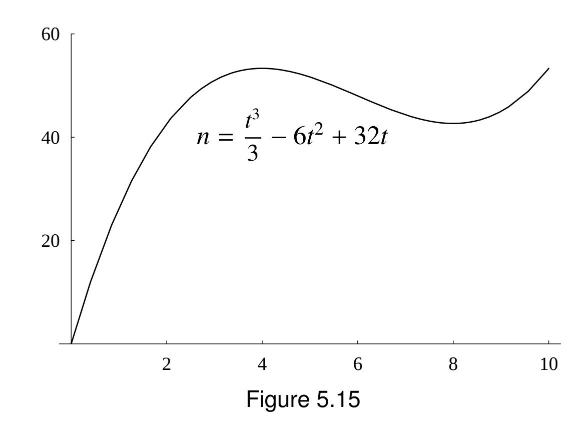


Figure 5.15

Example Find the dimensions of the rectangle that has maximum area if its perimeter is 20 cm.

Explanation The question asks for the length and width of the rectangle.
In the solution below, the domain of the area function A is not a closed interval. We can't use the steps as in the last example. Instead, we consider where A is increasing or decreasing.

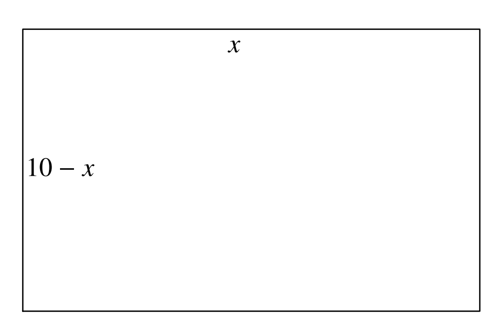


Figure 5.16

Solution Let the length of one side of the rectangle be x cm.
Then the length of an adjacent side is $(10 - x)$ cm.
Note that $0 < x$ and $0 < 10 - x$. Thus we have $0 < x < 10$.

The area A (in cm^2) of the rectangle is

$$A(x) = x(10 - x), \qquad 0 < x < 10.$$

We want to find the value of x at which A attains its maximum.

Differentiating $A(x)$, we get

$$\begin{aligned} A'(x) &= \frac{d}{dx}(10x - x^2) \\ &= 10 - 2x \quad (0 < x < 10). \end{aligned}$$

Solving $A'(x) = 0$, we obtain the critical number of A: $x_1 = 5$.

	$(0, 5)$	$(5, 10)$
A'	$+$	$-$
A	↗	↘

Since A is increasing on $(0, 5)$ and decreasing on $(5, 10)$, it follows that A attains its absolute maximum at $x_1 = 5$.

The dimensions of the largest rectangle is 5 cm × 5 cm. □

Remark The largest rectangle is, in fact, a square.

FAQ Can we include 0 and 10 in the domain of A?

Answer We may allow 0 and 10 in the domain of A. If $x = 0$ or 10, we get a rectangle one side of which is 0 cm. Such a figure is called a *degenerate rectangle*. Including the endpoints, the domain becomes a closed and bounded interval. Below we redo this problem using the method for the last example.

Alternative solution Let the length of one side of the rectangle be x cm. Then the length of an adjacent side is $(10 - x)$ cm. The area A (in cm^2) of the rectangle is

$$A(x) = x(10 - x), \qquad 0 \le x \le 10.$$

Differentiating $A(x)$, we get $A'(x) = 10 - 2x$ $(0 < x < 10)$.

Solving $A'(x) = 0$, we obtain the critical number of A: $x_1 = 5$. Comparing the values of A at the critical number and that at the endpoints:

x	0	5	10
$A(x)$	0	25	0

we see that A attains its maximum at $x_1 = 5$. Hence the dimensions of the largest rectangle is 5 cm × 5 cm. □

FAQ Can we apply the Second Derivative Test to check that A has maximum at $x_1 = 5$?

Answer If you use the Second Derivative Test, you can only tell that A has *local* maximum at $x_1 = 5$. In this problem, we want *global* maximum.

However, there is a special version of the Second Derivative Test which can be applied to this problem. □

Second Derivative Test (Special Version) *Let f be a function and let x_0 be a real number such that f is differentiable on an open interval (a, b) containing x_0. Suppose that x_0 is the only critical number of f in (a, b).*

(1) *If $f''(x_0) < 0$, then in (a, b), f attains its maximum at x_0, that is, $f(x_0) \geq f(x)$ for all $x \in (a, b)$.*

(2) *If $f''(x_0) > 0$, then in (a, b), f attains its minimum at x_0, that is, $f(x_0) \leq f(x)$ for all $x \in (a, b)$.*

Explanation Below we give a proof for (1). For this, we use a method called *Proof by Contradiction*. The result we want to prove is in the form "*Assumption*; *Conclusion*".

- The assumption is "*f is differentiable on an open interval (a, b) containing x_0 and x_0 is the only critical number of f in (a, b)*".
- The conclusion is "*If $f''(x_0) < 0$, then in (a, b), f attains its maximum at x_0*".

The negation (opposite) of the conclusion is "*It is not true that if $f''(x_0) < 0$, then in (a, b), f attains its maximum at x_0*" which can be restated as "*$f''(x_0) < 0$ and in (a, b), f does not attain its maximum at x_0*".

The method of *Proof by Contradiction* is to assume that the conclusion is false and use it (together with the given assumption) to deduce something that contradicts the given assumption. More specifically, we want to deduce that there exists $x_2 \in (a, b)$ with $x_2 \neq x_0$ such that $f'(x_2) = 0$, which contradicts the assumption that x_0 is the only critical number of f in (a, b).

In the proof below, we write "*Without loss of generality, we may assume that $x_1 > x_0$*". It means that the other case where $x_1 < x_0$ can be treated similarly.

Proof We give a proof for (1). For (2), it can be proved similarly or alternatively proved by applying (1) to the function $-f$.

Suppose that (1) does not hold, that is, suppose that $f''(x_0) < 0$ but there exists $x_1 \in (a, b)$ such that

$$f(x_1) > f(x_0).$$

Without loss of generality, we may assume that $x_1 > x_0$. Applying the Extreme Value Theorem to f on the interval $[x_0, x_1]$, we see that there exists $x_2 \in [x_0, x_1]$ such that

$$f(x_2) \leq f(x) \quad \text{for all } x \in [x_0, x_1].$$

It is clear that $x_2 \neq x_1$. Moreover, we have $x_2 \neq x_0$; this is because

$$f(x_0) > f(x) \quad \text{if } x \text{ is sufficiently close to } x_0 \text{ and } x \neq x_0$$

by the Second Derivative Test (since $f''(x_0) < 0$). Thus we have

$$x_2 \in (x_0, x_1) \quad \text{and} \quad f(x_2) \leq f(x) \text{ for all } x \in (x_0, x_1),$$

which implies that f has a local minimum at x_2. By Theorem 5.1.3, we have $f'(x_2) = 0$, which (together with that $x_2 \neq x_0$) contradicts the assumption that x_0 is the only critical number of f in (a, b). □

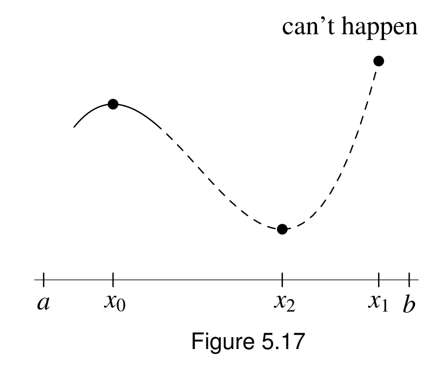


Figure 5.17

Alternative solution to the rectangle problem Let the length of one side of the rectangle be x cm. Then the length of an adjacent side is $(10 - x)$ cm. The area A (in cm^2) of the rectangle is

$$A(x) = x(10 - x), \qquad 0 < x < 10.$$

Differentiating $A(x)$, we get $A'(x) = 10 - 2x \ (0 < x < 10)$.

Solving $A'(x) = 0$, we obtain the critical number of A: $x_1 = 5$.

Differentiating $A'(x)$, we get $A''(x) = \frac{d}{dx}(10 - 2x) = -2$.

Since $A''(5) = -2 < 0$ and 5 is the only critical number of A in $(0, 10)$, it follows from the Second Derivative Test (Special Version) that A attains its maximum at $x_1 = 5$. Hence the dimensions of the largest rectangle is 5 cm × 5 cm. □

Example A rectangular box without lid is to be made from a square cardboard of sides 18 cm by cutting equal squares from each corner and then folding up the sides. Find the length of the side of the square that must be cut off if the volume of the box is to be maximized. What is the maximum volume?

Solution 1 Let the length of the side of the square to be cut off be x cm. Then the base of the box is a square with each side equals to $(18 - 2x)$ cm. Hence we have $0 < x < 9$.

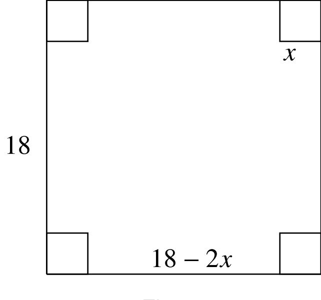


Figure 5.18

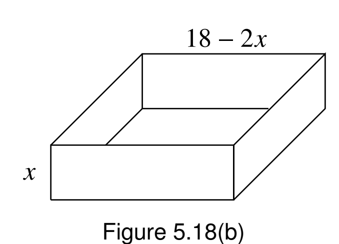


Figure 5.18(b)

The volume V, in cm^3, of the open box is

$$V(x) = x(18 - 2x)^2, \qquad 0 < x < 9.$$

Differentiating $V(x)$, we get

$$\begin{aligned} V'(x) &= \frac{d}{dx}(324x - 72x^2 + 4x^3) \\ &= 324 - 144x + 12x^2 \qquad (0 < x < 9) \end{aligned}$$

Solving $V'(x) = 0$, that is
$$\begin{aligned} 324 - 144x + 12x^2 &= 0 \quad (0 < x < 9) \\ 12(x-9)(x-3) &= 0 \quad (0 < x < 9) \end{aligned}$$
we get the critical number of V in $(0, 9)$: $x_1 = 3$.

	$(0,3)$	$(3,9)$
V'	$+$	$-$
V	$\nearrow$	$\searrow$

Since V is increasing on $(0, 3)$ and decreasing on $(3, 9)$, it follows that on $(0, 9)$, V attains its maximum at $x_1 = 3$.

To maximize the volume of the box, the length of the side of the square that must be cut off is 3 cm.

The maximum volume is $V(3) = 432\,\text{cm}^3$. □

Solution 2 Let the length of the side of the square to be cut off be x cm. Then the base of the box is a square with each side equals to $(18 - 2x)$ cm. Hence we have $0 \le x \le 9$ (when $x = 0$ or 9, we get a degenerate box with zero volume).

The volume V, in cm^3, of the open box is

$$V(x) = x\,(18 - 2x)^2, \qquad 0 \le x \le 9.$$

Differentiating $V(x)$, we get
$$\begin{aligned} V'(x) &= \frac{\mathrm{d}}{\mathrm{d}x}(324x - 72x^2 + 4x^3) \\ &= 324 - 144x + 12x^2 \qquad (0 < x < 9) \end{aligned}$$

Solving $V'(x) = 0$, that is
$$\begin{aligned} 324 - 144x + 12x^2 &= 0 \quad (0 < x < 9) \\ 12(x-9)(x-3) &= 0 \quad (0 < x < 9) \end{aligned}$$
we get the critical number of V in $(0, 9)$: $x_1 = 3$.

Comparing the value of V at the critical number and that the the endpoints:

x	0	3	9
$V(x)$	0	432	0

we see that to have maximum volume, the length of the side of the square that must be cut off is 3 cm; and that the maximum volume is $432\,\text{cm}^3$. □

Solution 3 Let the length of the side of the square to be cut off be x cm. Then the base of the box is a square with each side equals to $(18 - 2x)$ cm. Hence we have $0 < x < 9$.

The volume V, in cm^3, of the open box is

$$V(x) = x\,(18 - 2x)^2, \qquad 0 < x < 9.$$

Differentiating $V(x)$, we get
$$\begin{aligned} V'(x) &= \frac{\mathrm{d}}{\mathrm{d}x}(324x - 72x^2 + 4x^3) \\ &= 324 - 144x + 12x^2 \qquad (0 < x < 9) \end{aligned}$$

Solving $V'(x) = 0$, that is
$$\begin{aligned} 324 - 144x + 12x^2 &= 0 \quad (0 < x < 9) \\ 12(x-9)(x-3) &= 0 \quad (0 < x < 9) \end{aligned}$$
we get the critical number of V in $(0, 9)$: $x_1 = 3$.

Differentiating $V'(x)$, we get
$$\begin{aligned} V''(x) &= \frac{\mathrm{d}}{\mathrm{d}x}(324 - 144x + 12x^2) \\ &= -144 + 24x. \end{aligned}$$

Since $V''(3) = -72 < 0$ and $x_1 = 3$ is the only critical number of V in $(0, 9)$, it follows from the Second Derivative Test (Special Version) that in $(0, 9)$, V attains its maximum at $x_1 = 3$. Thus the length of the side of the square that must be cut off is 3 cm and the maximum volume is $V(3) = 432\,\text{cm}^3$. □

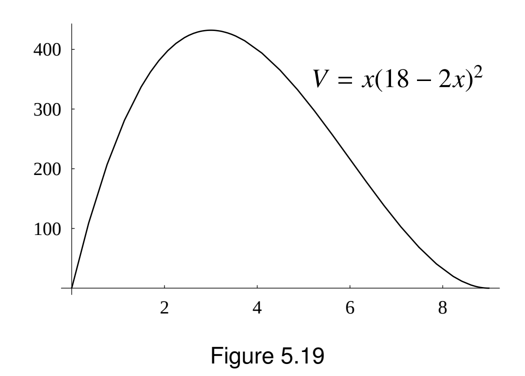


Figure 5.19

5.2.3 Applications to Economics

Suppose a manufacturer produces and sells a product. Denote $C(q)$ to be the total cost for producing and marketing q units of the product. Thus C is a function of q and it is called the *(total) cost function*. The rate of change of C with respect to q is called the *marginal cost*, that is,

$$\text{marginal cost} = \frac{dC}{dq}.$$

Denote $R(q)$ to be the total amount received for selling q units of the product. Thus R is a function of q and it is called the *revenue function*. The rate of change of R with respect to q is called the *marginal revenue*, that is,

$$\text{marginal revenue} = \frac{dR}{dq}.$$

Denote $P(q)$ to be the profit of producing and selling q units of the product, that is,

$$P(q) = R(q) - C(q).$$

Thus P is a function of q and it is called the *profit function*.

Denote q_{max} to be the largest number of units of the product that the manufacturer can produce. Assuming that q can take any value between 0 and q_{max}. Then for each of the functions C, R and P, the domain is $[0, q_{max}]$. Suppose that the cost function and the revenue function are differentiable on $(0, q_{max})$ and suppose that producing 0 or q_{max} units of the product will not give maximum profit. Then in order to have maximum profit, we need

$$\frac{dP}{dq} = 0,$$

or equivalently,

$$\frac{dC}{dq} = \frac{dR}{dq},$$

that is, marginal cost = marginal revenue.

Example The demand equation for a certain product is

$$q - 90 + 2p = 0, \qquad 0 \le q \le 90,$$

where q is the number of units and p is the price per unit, and the average cost function is

$$C_{\text{av}} = q^2 - 8q + 57 + \frac{2}{q} \qquad 0 < q \leq 90.$$

At what value of q will there be maximum profit? What is the maximum profit?

Explanation Although the average cost function is undefined at $q = 0$, we may include 0 in the domain of the cost function. The cost function and the revenue function are differentiable on $(0, 90)$. However, we do not know whether maximum profit would be attained in $(0, 90)$ or at an endpoint. So we use the method for finding absolute extrema for functions on closed and bounded intervals.

Solution The cost function C is given by

$$C(q) = q \cdot C_{\text{av}} = q^3 - 8q^2 + 57q + 2 \qquad (0 \leq q \leq 90),$$

and the revenue function R is given by

$$R(q) = p \cdot q = \frac{90 - q}{2} \cdot q \qquad (0 \leq q \leq 90).$$

Therefore the profit function P is given by

$$\begin{aligned} P(q) &= R(q) - C(q) \\ &= \left(45q - \frac{q^2}{2}\right) - (q^3 - 8q^2 + 57q + 2) \\ &= -q^3 + \frac{15}{2}q^2 - 12q - 2, \qquad (0 \leq q \leq 90). \end{aligned}$$

Differentiating $P(q)$, we get
$$\begin{aligned} P'(q) &= \frac{\mathrm{d}}{\mathrm{d}q}\left(-q^3 + \frac{15}{2}q^2 - 12q - 2\right) \\ &= -3q^2 + 15q - 12 \qquad (0 < q < 90) \end{aligned}$$

Solving $P'(q) = 0$, that is,
$$\begin{aligned} -3q^2 + 15q - 12 &= 0 \quad (0 < q < 90) \\ -3(q-1)(q-4) &= 0 \quad (0 < q < 90), \end{aligned}$$

we get the critical numbers of P: $q_1 = 1$ and $q_2 = 4$.

Comparing the values of P at the critical numbers as well as that at the endpoints:

q	0	1	4	90
$P(q)$	-2	$-\frac{15}{2}$	6	-669332

we see that maximum profit is attained at $q_2 = 4$ and the maximum profit is 6 (units of money). □

Remark If we know that maximum profit is not attained at the endpoints, we can simply compare the values of P at q_1 and q_2.

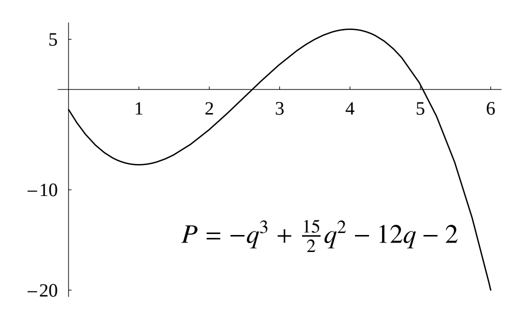


Figure 5.20

Exercise 5.2

1. For each of the given function f, find its absolute extrema on the given interval.
   (a) $f(x) = 4x^3 + 3x^2 - 18x + 1, \qquad [0, 3]$
   (b) $f(x) = -3x^5 + 5x^3 + 2, \qquad [-2, 0]$
   (c) $f(x) = 1 + 2x^3 - 3x^4, \qquad [-1, 1]$
2. Find two positive real numbers whose sum is 50 and whose product is a maximum.
3. Find two real numbers x and y satisfying $2x + y = 15$ such that $x^2 + y^2$ is minimized.
   Can you find a geometric meaning for the result?
4. Find the dimensions of the rectangle of area 100 square units that has the least perimeter.
5. A rectangular field is to be enclosed by a fence and divided equally into two parts by a fence parallel to one pair of the sides. If a total of 600 m of fence is to be used, find the dimensions of the field if its area is to be maximized.
6. A book is to contain $36\,\text{in}^2$ of printed matter per page, with margins of 1 in along the sides and $1\frac{1}{2}$ in along the top and bottom. Find the dimensions of the page that will require the minimum amount of paper.
7. Suppose that a ball is thrown straight up into the air and its height after t seconds is $5 + 24t - 16t^2$ feet. Determine how long it will take the ball to reach its maximum height and determine the maximum height.
8. It is known from experiments that the height (in meter) of a certain plant after t months is given (approximately) by
   $$h(t) = \sqrt{t} - t, \qquad 0 \le t \le 1.$$
   How long, on the average, will it take a plant to reach its maximum height? What is the maximum height?
9. A company manufactures and sells x pieces of a certain product per month. The monthly cost (in dollars) is
   $$C(x) = 120000 + 100x$$
   and the price-demand equation is
   $$p = 300 - \frac{x}{15}$$
   where $0 \le x \le 4000$. Find the maximum profit, the production level that will give the maximum profit, and the price the company should charge for each piece of the product.

Chapter 6

Integration

6.1 Definite Integrals

In the introduction of Chapter 3, we consider the area of the region under the curve $y = x^2$ and above the x-axis for x between 0 and 1. To get approximations for the area, we divide $[0, 1]$ into n equal subintervals:

$$[x_0, x_1], \quad [x_1, x_2], \quad \dots, \quad [x_{n-1}, x_n],$$

where $x_i = \frac{i}{n}$ for $0 \le i \le n$; and for each $i = 1, \dots, n$, in the subinterval $[x_{i-1}, x_i]$, we take the left endpoint x_{i-1} and consider the sum $\sum_{i=1}^{n} f(x_{i-1}) \cdot \frac{1}{n}$, that is, $f(x_0) \cdot \frac{1}{n} + \cdots + f(x_{n-1}) \cdot \frac{1}{n}$. We have seen that the sum is close to $\frac{1}{3}$ (which is the required area) if n is large. Using limit notation, the result can be written as

$$\lim_{n\to\infty} \sum_{i=1}^{n} f(x_{i-1}) \cdot \frac{1}{n} = \frac{1}{3}. \tag{6.1.1}$$

The above idea can be generalized to any continuous functions f on any closed and bounded interval. Moreover, f need not be non-negative.

Theorem 6.1.1 *Let f be a function that is continuous on a closed and bounded interval $[a, b]$. Then the following limit exists:*

$$\lim_{n\to\infty} \sum_{i=1}^{n} f(x_{i-1}) \cdot \frac{b-a}{n},$$

where $x_i = a + \frac{i}{n}(b - a)$ for $0 \le i \le n$.

Explanation

- By the construction of the x_i's, we have $x_0 = a$, $x_n = b$, $x_0 < x_1 < \cdots < x_n$, and for every $i = 1, \dots, n$, the subinterval $[x_{i-1}, x_i]$ has length $\frac{b-a}{n}$ and x_{i-1} is the left-endpoint of the subinterval.
- If f is non-negative on $[a, b]$, that is, $f(x) \ge 0$ for all $x \in [a, b]$, then $\lim_{n\to\infty} \sum_{i=1}^{n} f(x_{i-1}) \cdot \frac{b-a}{n}$ is the area bounded by the graph of f, the x-axis and the vertical lines given by $x = a$ and $x = b$.

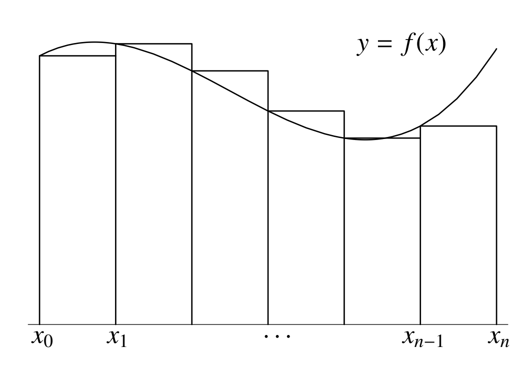


Figure 6.1

Definition Let f be a function that is continuous on a closed and bounded interval $[a, b]$. The number $\lim_{n\to\infty} \sum_{i=1}^{n} f(x_{i-1}) \cdot \frac{b-a}{n}$, where $x_i = a + \frac{i}{n}(b-a)$ for $0 \le i \le n$, is called the *definite integral* of f *from* a *to* b and is denoted by $\int_a^b f(x)\,\mathrm{d}x$, that is,

$$\int_a^b f(x)\,\mathrm{d}x = \lim_{n\to\infty} \sum_{i=1}^{n} f(x_{i-1}) \cdot \frac{b-a}{n}. \tag{6.1.2}$$

Example The result given in (6.1.1) can be written as

$$\int_0^1 x^2\,\mathrm{d}x = \frac{1}{3}. \tag{6.1.3}$$

Remark In Theorem 6.1.1, in each subinterval $[x_{i-1}, x_i]$, instead of taking the left endpoint x_{i-1}, we can take the right endpoint x_i (see Figure 6.2) and we have

$$\lim_{n\to\infty} \sum_{i=1}^{n} f(x_i) \cdot \frac{b-a}{n} = \int_a^b f(x)\,\mathrm{d}x. \tag{6.1.4}$$

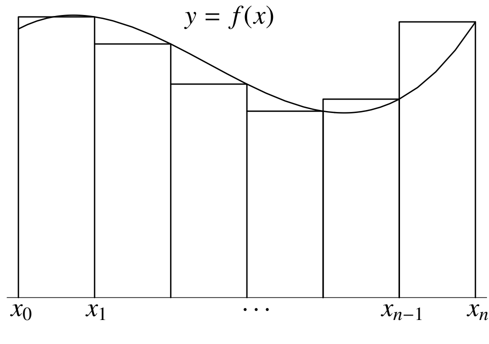


Figure 6.2

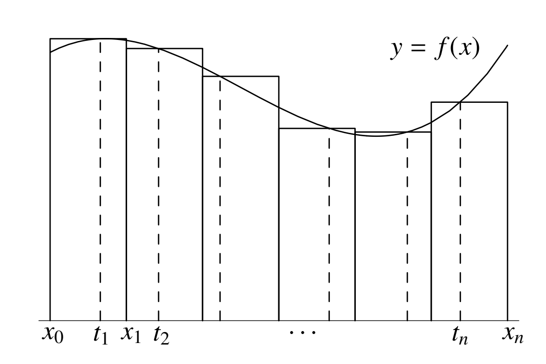


Figure 6.3

In fact, we can take an arbitrarily point (denoted by t_i) in $[x_{i-1}, x_i]$: the sum $\sum_{i=1}^{n} f(t_i) \cdot \frac{b-a}{n}$ is close to $\int_a^b f(x)\,\mathrm{d}x$ if n is large enough (see Figure 6.3).

More generally, the subintervals $[x_0, x_1], \ldots, [x_{n-1}, x_n]$ need not be of equal lengths. All we need is that the lengths are small enough: if $a = x_0 < x_1 < \cdots < x_n = b$ and $\Delta x_1, \ldots, \Delta x_n$ are small enough, where Δx_i is the length of the ith subinterval $[x_{i-1}, x_i]$, then for every choice of $t_1, \ldots, t_n$ with $t_i \in [x_{i-1}, x_i]$ for $1 \le i \le n$, the sum (called a *Riemann Sum*)

$$\sum_{i=1}^{n} f(t_i)\Delta x_i$$

is close to $\int_a^b f(x)\,\mathrm{d}x$. Many authors use this to define definite integral.

Below, we apply (6.1.4) to deduce the result given in (6.1.3). For this, we take $f(x) = x^2$, $a = 0$, $b = 1$ and $x_i = \frac{i}{n}$ for $0 \le i \le n$.

Example By (6.1.4), we have

$$\begin{aligned}
\int_0^1 x^2\,\mathrm{d}x &= \lim_{n\to\infty}\sum_{i=1}^{n}\left(\frac{i}{n}\right)^2\cdot\frac{1}{n} \\
&= \lim_{n\to\infty}\frac{1}{n^3}\sum_{i=1}^{n} i^2 \\
&= \lim_{n\to\infty}\frac{1}{n^3}\cdot\frac{n(n+1)(2n+1)}{6} && \text{Sum of Squares Formula} \\
&= \lim_{n\to\infty}\frac{2n^3}{6n^3} && \text{Leading Term Rule} \\
&= \frac{1}{3}
\end{aligned}$$

FAQ Can we define $\int_a^b f(x)\,\mathrm{d}x$ if f is not continuous on $[a,b]$?

Answer In defining $\int_a^b f(x)\,\mathrm{d}x$, we need Theorem 6.1.1. The condition "*f is continuous on* $[a,b]$" is used to guarantee that $\lim\limits_{n\to\infty}\sum\limits_{i=1}^{n} f(x_{i-1})\cdot\frac{1}{n}$ exists.

In general, if f is a function defined on $[a,b]$ such that there exists a (unique) real number I satisfying

(∗) $\sum\limits_{i=1}^{n} f(t_i)\Delta x_i$ is arbitrarily close I if $\Delta x_1,\ldots,\Delta x_n$ are sufficiently small, where $\Delta x_i = x_i - x_{i-1}$ for $1\le i\le n$, $a = x_0 < x_1 < \cdots < x_n = b$ and $t_i \in [x_{i-1}, x_i]$ for $1\le i\le n$,

then the unique number I is defined to be $\int_a^b f(x)\,\mathrm{d}x$. □

Remark

- In view of (∗), we may write
$$\int_a^b f(x)\,\mathrm{d}x = \lim_{\|\Delta\|\to 0}\sum_{i=1}^{n} f(t_i)\Delta x_i,$$
where $\|\Delta\|\to 0$ means the lengths Δx_i's tend to zero. However, this kind of limit is different from that discussed in Chapter 3.
- The symbol $\int$ was introduced by Leibniz and is called the *integral sign*. It is an elongated S and was chosen because a definite integral is a limit of sums.
- Since the definite integral of a (continuous) function on $[a,b]$ depends on the function f and the interval $[a,b]$ only, it can simply be denoted by $\int_a^b f$, omitting the variable x and the notation dx. However, the notation $\int_a^b f(x)\,\mathrm{d}x$ is preferred. There are two reasons:
  (1) the notation dx reminds us of the factors Δx_i in the sums $\sum\limits_{i=1}^{n} f(t_i)\Delta x_i$;
  (2) with the variable included in the notation, it is easier to handle the *substitution method* for integration (see Chapter 10).
- In the notation $\int_a^b f(x)\,\mathrm{d}x$, the variable x is called a *dummy variable*; it can be replaced by any other symbol. For example, using t as the dummy variable, (6.1.3) can be written as $\int_0^1 t^2\,\mathrm{d}t = \frac{1}{3}$. Note that if we use t as the dummy variable, we have to change dx to dt accordingly.

Example Use definition to find $\displaystyle\int_1^2 x\,\mathrm{d}x$.

Solution Applying (6.1.2) to $f(x) = x$, $a = 1$, $b = 2$ and $x_i = 1 + \frac{i}{n}$ for $0 \le i \le n$, we get

$$\begin{aligned}
\int_1^2 x\,\mathrm{d}x &= \lim_{n\to\infty} \sum_{i=1}^{n} \left(1 + \frac{i-1}{n}\right) \cdot \frac{1}{n} \\
&= \lim_{n\to\infty} \frac{1}{n^2} \sum_{i=1}^{n} (n + i - 1) \\
&= \lim_{n\to\infty} \frac{1}{n^2} \cdot \frac{n(n + 2n - 1)}{2} && \text{Sum of A.P.} \\
&= \lim_{n\to\infty} \frac{3n - 1}{2n} \\
&= \lim_{n\to\infty} \frac{3n}{2n} && \text{Leading Term Rule} \\
&= \frac{3}{2}.
\end{aligned}$$

□

Remark The value of the definite integral is the area of the trapezoidal region shown in Figure 6.4. Readers can check that the result agrees with that obtained by using formula for area of trapezoid.

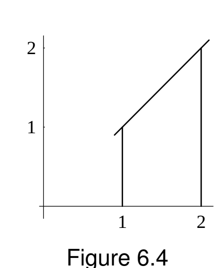


Figure 6.4

Definite Integral for Constant Functions Let c be a constant and let a and b be real numbers with $a < b$. Then we have $\displaystyle\int_a^b c\,\mathrm{d}x = c \cdot (b - a)$.

Proof Applying (6.1.2) to $f(x) = c$ and $x_i = a + \frac{i}{n}(b - a)$ for $0 \le i \le n$, we get

$$\begin{aligned}
\int_a^b c\,\mathrm{d}x &= \lim_{n\to\infty} \sum_{i=1}^{n} c \cdot \frac{b-a}{n} \\
&= \lim_{n\to\infty} c \cdot \frac{b-a}{n} \times n && \text{Sum of Constants} \\
&= \lim_{n\to\infty} c \cdot (b - a) && \text{Rule (L1) for Limit} \\
&= c \cdot (b - a)
\end{aligned}$$

□

Remark If $c > 0$, then $\int_a^b c\,\mathrm{d}x$ is the area of the rectangular region shown in Figure 6.5.

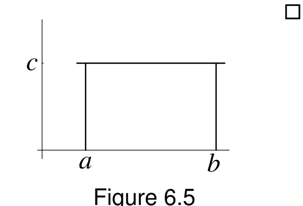


Figure 6.5

Rules for Definite Integrals Let f and g be functions that are continuous on a closed and bounded interval $[a, b]$. Let α be a constant and let $c \in (a, b)$. Then we have

(Int1) $\displaystyle\int_a^b \big[f(x) + g(x)\big]\,\mathrm{d}x = \int_a^b f(x)\,\mathrm{d}x + \int_a^b g(x)\,\mathrm{d}x$

Proof Apply definition and Rule (L4) for limits of sequences. □

(Int2) $\displaystyle\int_a^b \alpha f(x)\,\mathrm{d}x = \alpha \int_a^b f(x)\,\mathrm{d}x$

Proof Apply definition and Rule (L5s) for limits of sequences. □

Remark Using Rules (Int1) and (Int2), we get

$$\int_a^b \big[f(x) - g(x)\big]\,\mathrm{d}x = \int_a^b f(x)\,\mathrm{d}x - \int_a^b g(x)\,\mathrm{d}x.$$

In fact, Rule (Int1) is valid for sum and difference of finitely many (continuous) functions.

(Int3) $\displaystyle\int_a^b f(x)\,\mathrm{d}x = \int_a^c f(x)\,\mathrm{d}x + \int_c^b f(x)\,\mathrm{d}x.$

Explanation The proof for the result is not easy. For the case where f is nonnegative on $[a, b]$, the result can be seen from the geometric interpretation shown in Figure 6.6. □

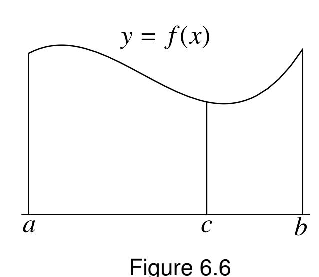


Figure 6.6

Example
$$\begin{aligned}\int_0^1 5x^2\,\mathrm{d}x &= 5\int_0^1 x^2\,\mathrm{d}x && \text{Rule (Int2)}\\ &= 5 \times \frac{1}{3} && \text{by (6.1.3)}\\ &= \frac{5}{3}\end{aligned}$$

Example
$$\begin{aligned}\int_1^2 (3 - x)\,\mathrm{d}x &= \int_1^2 3\,\mathrm{d}x - \int_1^2 x\,\mathrm{d}x && \text{Rule (Int1)}\\ &= 3 \times (2 - 1) - \frac{3}{2} && \text{Definite Integral for Constant \& Example on Page 160}\\ &= \frac{3}{2}\end{aligned}$$

In defining $\int_a^b f(x)\,\mathrm{d}x$, we need $a < b$. For convenience, we introduce the following:

Convention Let f be a function that is continuous on a closed and bounded interval $[a, b]$ and let $c \in [a, b]$. Then we define

(1) $\displaystyle\int_b^a f(x)\,\mathrm{d}x = -\int_a^b f(x)\,\mathrm{d}x$

(2) $\displaystyle\int_c^c f(x)\,\mathrm{d}x = 0$

Example $\displaystyle\int_2^2 (1 + 2x - 3x^2)\,\mathrm{d}x = 0$ since the function $1 + 2x - 3x^2$ is continuous on $[0, 3]$ (for example) and $2 \in [0, 3]$.

Example
$$\int_1^0 x^2\,\mathrm{d}x = -\int_0^1 x^2\,\mathrm{d}x \quad \text{by convention}$$
$$= -\frac{1}{3} \quad \text{by (6.1.3)}$$

Terminology In a definite integral $\int_a^b f(x)\,\mathrm{d}x$,

- the function f is called the *integrand*;
- the numbers a and b are called the *limits of integration*; a is the *lower limit* and b the *upper limit*.

Exercise 6.1

1. For each of the following definite integrals, use the results in this section to find its value:

   (a) $\int_0^1 (1 - 3x^2)\,\mathrm{d}x$ (b) $\int_2^1 4x\,\mathrm{d}x$

2. Use definition to find the definite integral $\int_0^1 x^3\,\mathrm{d}x$.

   Given: $1^3 + 2^3 + \cdots + n^3 = \frac{n^2(n+1)^2}{4}$

6.2 Fundamental Theorem of Calculus

In the last section, we give the definition and some examples of definite integral. Although we have some rules that are useful in calculating definite integrals, we still have to know the definite integrals of some "*basic functions*". For example, using rules, we can find the definite integrals of polynomials provided that we know $\int_a^b x^n\,\mathrm{d}x$ for positive integers n. Finding definite integrals by first principle (that is, by definition) is very tedious. In this section, we describe a simple way (Fundamental Theorem of Calculus, Version 2) to find definite integrals. It is quite surprising that differentiation and integration are related— they are reverse process of each other (see Fundamental Theorem of Calculus, Versions 1 and 3).

Given a function f that is continuous on a closed and bounded interval $[a, b]$, in order to "find" the definite integral $\int_a^b f(x)\,\mathrm{d}x$, we introduce an auxiliary function F from $[a, b]$ into $\mathbb{R}$ defined by

$$F(x) = \int_a^x f(t)\,\mathrm{d}t \qquad a \leq x \leq b.$$

In the construction of F, for each $x \in (a, b]$, the value $F(x)$ is defined to be the definite integral of f over the interval $[a, x]$ and for $x = a$, by convention, $F(a) = \int_a^a f(t)\,\mathrm{d}t$ is defined to be 0. Note that x is used as the independent variable for the function F. For clarity, we use another symbol t as the dummy variable for the definite integral of f over $[a, x]$.

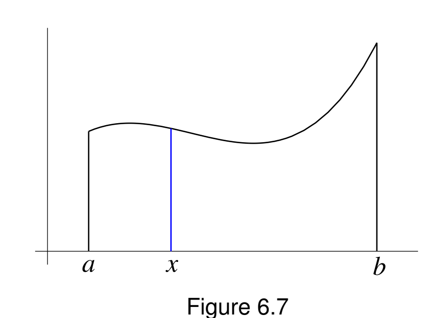


Figure 6.7

Geometrically, if f is nonnegative on $[a, b]$, then F can be considered as an "area function" with $F(x)$ equal to the area under the graph of f (and above the horizontal axis) from a to x.

By the construction of F, the required definite integral is $F(b)$. If we can find a formula for $F(x)$, then we can solve the problem. The following result gives a relation between F and f.

Fundamental Theorem of Calculus, Version 1 *Let f be a function that is continuous on a closed and bounded interval $[a, b]$. Let F be the function from $[a, b]$ into $\mathbb{R}$ defined by*

$$F(x) = \int_a^x f(t)\,\mathrm{d}t \qquad \text{for } a \le t \le b.$$

Then F is continuous on $[a, b]$ and differentiable on (a, b) with $F'(x) = f(x)$ for all $x \in (a, b)$.

Explanation The proof of this result will be given in the appendix. Below we explain how to "obtain" $F' = f$ on (a, b) intuitively for the case where f is nonnegative on $[a, b]$. Recall that

$$F'(x) = \lim_{h \to 0} \frac{F(x + h) - F(x)}{h}.$$

For $x \in (a, b)$ and for sufficiently small $h > 0$ (such that $a + h \le b$), we have

$$\begin{aligned} F(x+h) - F(x) &= \int_a^{x+h} f(t)\,\mathrm{d}t - \int_a^x f(t)\,\mathrm{d}t && \text{by construction of } F \\ &= \left(\int_a^x f(t)\,\mathrm{d}t + \int_x^{x+h} f(t)\,\mathrm{d}t\right) - \int_a^x f(t)\,\mathrm{d}t && \text{Rule (Int3)} \\ &= \int_x^{x+h} f(t)\,\mathrm{d}t. \end{aligned}$$

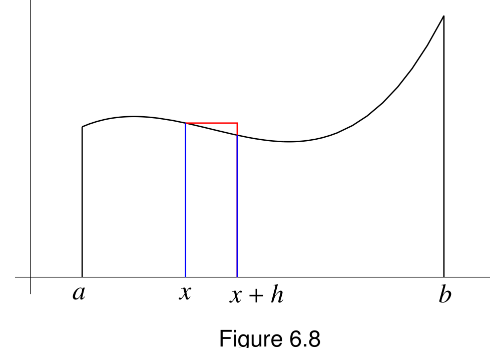


Figure 6.8

Note that $\int_x^{x+h} f(t)\,\mathrm{d}t$ is the area of the region below the graph of f (and above the horizontal axis) from x to $x + h$. If h is small, then $[x, x+h]$ is a short interval and the area of the small region under consideration can be approximated by the area of the rectangular region with base $[x, x+h]$ on the horizontal axis and height equal to $f(x)$. Thus we have

$$\int_x^{x+h} f(t)\,\mathrm{d}t \ \text{ is close to } \ f(x) \cdot h \ \text{ if } h \text{ is small,}$$

from which we obtain

$$\frac{F(x+h) - F(x)}{h} = \frac{\int_x^{x+h} f(t)\,\mathrm{d}t}{h} \quad \text{is close to } f(x) \ \text{ if } h \text{ is small.}$$

Taking limit, we get $F'(x) = f(x)$.

Remark To be more precise, the above argument gives $\lim\limits_{h \to 0+} \frac{F(x+h) - F(x)}{h} = f(x)$ only.

In view of the Fundamental Theorem of Calculus (Version 1), to find $\int_a^b f(x)\,\mathrm{d}x$, we should look for functions G such that $G' = f$.

Definition Let f be a function that is continuous on a closed and bounded interval $[a, b]$. Suppose that G is a function that is defined on $[a, b]$ such that the following two conditions are satisfied:

(1) G is continuous on $[a, b]$;

(2) G is differentiable on (a, b) and $G'(x) = f(x)$ for all $x \in (a, b)$.

Then we say that G is a *primitive* for f on $[a, b]$.

Example Let $f(x) = 3x^2$ and let $G(x) = x^3$. Note that f and G are continuous on $\mathbb{R}$ and that $G'(x) = f(x)$ for all $x \in \mathbb{R}$. Thus G is a primitive for f on every closed and bounded interval $[a, b]$.

Remark Primitive is not unique. For example, the function $G_1(x) = x^3 + 1$ is also a primitive for f (on every closed and bounded interval). In fact, for every constant C, the function

$$x^3 + C \tag{6.2.1}$$

is a primitive for f (on every closed and bounded interval). It is natural to ask whether there are any more primitives:

- If $F'(x) = f(x)$ for all $x \in \mathbb{R}$, must F be in the form (6.2.1)?

Corollary 6.2.2, which is based on the following theorem, tells that the answer is affirmative.

Theorem 6.2.1 *Let F and G be functions that are defined on a closed and bounded interval $[a, b]$. Suppose that F and G are continuous on $[a, b]$ and are differentiable on (a, b) with $F'(x) = G'(x)$ for all $x \in (a, b)$. Then on $[a, b]$, the functions F and G differ by a constant, that is, there exists a constant C such that*

$$F(x) - G(x) = C \qquad \text{for all } x \in [a, b].$$

Explanation The following is the geometry meaning of the result:

- The condition "$F'(x) = G'(x)$ *for all* $x \in (a, b)$" means that at corresponding points (same x-coordinates), tangents to the graphs of F and G are parallel.
- The conclusion is that the graph of F can be obtained from that of G by moving it upward ($C > 0$) or downward ($C < 0$).

Proof Let f be the function from $[a, b]$ into $\mathbb{R}$ defined by

$$f(x) = F(x) - G(x).$$

Note that $f'(x) = F'(x) - G'(x) = 0$ for all $x \in (a, b)$. Hence by Theorem 5.1.1, there exists a constant C such that

$$f(x) = C \qquad \text{for all } x \in (a, b).$$

Since f is continuous on $[a, b]$, it follows that

$$f(x) = C \qquad \text{for all } x \in [a, b]$$

from which we get the required result. □

Corollary 6.2.2 *Let f be a function that is continuous on a closed and bounded interval $[a, b]$. Suppose that F and G are primitives for f on $[a, b]$. Then on $[a, b]$, the functions F and G differ by a constant.*

Proof This is an immediate consequence of Theorem 6.2.1 since by the definition of primitive, the functions F and G are continuous on $[a, b]$ and differentiable on (a, b) and $F'(x) = f(x) = G'(x)$ for all $x \in (a, b)$. □

Example Find the value of the definite integral $\int_1^2 3x^2 \, dx$.

Solution Let $f(x) = 3x^2$ and let $G(x) = x^3$. Then G is a primitive for f on $[1, 2]$.

By the Fundamental Theorem of Calculus (Version 1), the function F given by

$$F(x) = \int_1^x 3t^2, dt, \qquad 1 \leq x \leq 2.$$

is a primitive for f on $[1, 2]$.

By Corollary 6.2.2, on the interval $[1, 2]$, the functions F and G differ by a constant, that is, there exists a constant C such that

$$F(x) - x^3 = C \qquad \text{for all } x \in [1, 2].$$

Putting $x = 1$ and using the construction of F, we get

$$0 - 1 = C.$$

which implies that $F(x) = x^3 - 1$ for all $x \in [1, 2]$. Hence, we have

$$\begin{aligned} \int_1^2 3x^2 \, dx &= F(2) \qquad \text{by construction of } F \\ &= 2^3 - 1 \\ &= 7. \end{aligned}$$

□

Remark The above procedure can be used to find the definite integral of f on any closed and bounded interval $[a, b]$. This is because G is a primitive for f on every $[a, b]$.

From the above example, we see that given a function f that is continuous on a closed and bounded interval $[a, b]$, if we can find a primitive for f over $[a, b]$, then we can find the definite integral $\int_a^b f(x) \, dx$. The following result describe an alternative procedure for finding $\int_a^b f(x) \, dx$ (there is no need to find the constant C).

Fundamental Theorem of Calculus, Version 2 *Let f be a function that is continuous on a closed and bounded interval $[a, b]$. Suppose that G is a primitive for f on $[a, b]$. Then we have*

$$\int_a^b f(x) \, dx = G(b) - G(a).$$

Proof Let F be the function from $[a, b]$ into $\mathbb{R}$ defined by

$$F(x) = \int_a^x f(t) \, dt, \qquad a \leq x \leq b.$$

By the Fundamental Theorem of Calculus, Version 1, the function F is a primitive for f on $[a, b]$. Hence by Corollary 6.2.2, there exists a constant C such that

$$F(x) - G(x) = C \quad \text{for all } x \in [a, b]. \tag{6.2.2}$$

Therefore, we have

$$\begin{aligned}
\int_a^b f(x)\,\mathrm{d}x &= F(b) && \text{by construction of } F \\
&= F(b) - F(a) && \text{since } F(a) = 0 \\
&= \big(G(b) + C\big) - \big(G(a) + C\big) && \text{by (6.2.2)} \\
&= G(b) - G(a)
\end{aligned}$$

□

Below we redo the last example using the second version of the Fundamental Theorem of Calculus.

Example Evaluate $\displaystyle\int_1^2 3x^2\,\mathrm{d}x$

Solution Since the function $G(x) = x^3$ is a primitive for the function $3x^2$ on the interval $[1, 2]$, it follows from the Fundamental Theorem of Calculus (Version 2) that

$$\begin{aligned}
\int_1^2 3x^2\,\mathrm{d}x &= G(2) - G(1) \\
&= 2^3 - 1^3 \\
&= 7.
\end{aligned}$$

□

FAQ Can we use other primitives for f?

Answer The Fundamental Theorem tells that any primitive will work. *Try it yourselves.* □

Notation We will use the notation $G(b) - G(a)$ quite often. For simplicity, it will be denoted by

$$\big[G(x)\big]_a^b \quad \text{or} \quad G(x)\Big|_a^b.$$

Example Find $\displaystyle\int_3^5 2x\,\mathrm{d}x.$

Explanation To find $\int_a^b f(x)\,\mathrm{d}x$, in applying the Fundamental Theorem of Calculus (Version 2), we have to find a function G that is continuous on $[a, b]$ such that $G' = f$ on (a, b). In this course, functions that we considered are "nice"—there is no need to check continuity; we just need to check that $G' = f$ (usually valid on a much larger interval).

Solution By inspection, we see that the function x^2 is a primitive for the integrand $2x$ (on every closed and bounded interval). Thus by the Fundamental Theorem of Calculus (Version 2), we have

$$\begin{aligned}
\int_3^5 2x\,\mathrm{d}x &= \big[x^2\big]_3^5 \\
&= 5^2 - 3^2 \\
&= 16.
\end{aligned}$$

□

Remark The definite integral is the area of the trapezoidal region that lies below the line $y = 2x$, above the x-axis and is bounded on the left and right by the vertical lines $x = 3$ and $x = 5$. *Use formula to check the answer yourselves.*

Exercise 6.2

1. For each of the following functions f, use inspection to find a primitive. Is your answer a primitive for f on every closed and bounded interval $[a, b]$? If not, what can you tell about $[a, b]$?

   (a) $f(x) = x$ (b) $f(x) = 1$

   (c) $f(x) = x^5$ (d) $f(x) = 2x + 1$

   (e) $f(x) = \frac{1}{2\sqrt{x}}$ (f) $f(x) = \sqrt{x}$

2. Use inspection to find a primitive for $f(x) = x^4$ and hence evaluate the following definite integrals.

   (a) $\int_0^1 x^4 \, dx$ (b) $\int_1^3 x^4 \, dx$

   (c) $\int_0^3 x^4 \, dx$ (d) $\int_0^3 7x^4 \, dx$

6.3 Indefinite Integrals

Suppose that f and F are functions that are continuous on a closed and bounded interval $[a, b]$. To show that F is a primitive for f on $[a, b]$ means to show that $F'(x) = f(x)$ for all $x \in (a, b)$. In view of this, we introduce a concept similar to *primitive*.

Definition Let f be a function that is continuous on an open interval (a, b). Suppose F is a function defined on (a, b) such that $F'(x) = f(x)$ for all $x \in (a, b)$. Then we say that F is an *antiderivative* for f on (a, b).

Example

(1) Let $f(x) = x^2$ and let $F(x) = \frac{1}{3}x^3$. Then we have

$$F'(x) = f(x) \quad \text{for all } x \in \mathbb{R}.$$

Thus F is an antiderivative for f on every open interval contained in $\mathbb{R}$.

(2) Let $g(x) = \frac{1}{\sqrt{x}}$ and let $G(x) = 2\sqrt{x}$. Then we have

$$G'(x) = g(x) \quad \text{for all } x > 0.$$

Thus G is an antiderivative for g on every open interval contained in $(0, \infty)$.

Remark Suppose that F is an antiderivative for f on (a, b). Then F is a primitive for f on every closed and bounded interval $[c, d]$ contained in (a, b). If in addition, F and f are defined at a and b and are continuous on $[a, b]$, then F is a primitive for f on $[a, b]$.

The following result is similar to that given in Corollary 6.2.2.

Theorem 6.3.1 *Let f be a function that is continuous on an open interval (a, b). Suppose that F and G are antiderivatives for f on (a, b). Then on (a, b), the functions F and G differ by a constant.*

Proof Apply the proof (the first part) for Theorem 6.2.1 □

Theorem 6.3.1 means that if we can find one antiderivative for a continuous function f on an open interval (a, b), then we can find all. More precisely, if F is an antiderivative for f on (a, b), then all the antiderivatives for f on (a, b) are in the form

$$F(x) + C, \qquad a < x < b \tag{6.3.1}$$

where C is a constant.

Note that (6.3.1) represents a family of functions defined on (a, b)—there are infinitely many of them, with each C corresponds to an antiderivative for f and vice versa. We call the family to be the *indefinite integral* of f (with respect to x) and we denote it by

$$\int f(x)\,\mathrm{d}x.$$

That is,

$$\int f(x)\,\mathrm{d}x = F(x) + C, \qquad a < x < b,$$

where F is a function such that $F'(x) = f(x)$ for all $x \in (a, b)$ and C is an arbitrary constant, called *constant of integration*.

Example Using the two results in the last example, we have the following:

(1) $\displaystyle\int x^2\,\mathrm{d}x = \frac{1}{3}x^3 + C, \quad -\infty < x < \infty,$ where C is an arbitrary constant.

(2) $\displaystyle\int \frac{1}{\sqrt{x}}\,\mathrm{d}x = 2\sqrt{x} + C, \quad x > 0,$ where C is an arbitrary constant.

Remark

- Sometimes, for simplicity, we write $\int x^2\,\mathrm{d}x = \frac{1}{3}x^3 + C$ etc.
  - ◊ The interval $\mathbb{R}$ is omitted because it can be determined easily.
  - ◊ The symbol C is understood to be an arbitrary constant.
- Since we can use any symbol to denote the independent variable, we may also write $\int t^2\,\mathrm{d}t = \frac{1}{3}t^3 + C$ etc.
- Instead of a family of functions, sometimes we write $\int f(x)\,\mathrm{d}x$ to represent a function only. See the discussion in the *Alternative Solution* on page 177.

Terminology

- To *integrate* a function f means to find the indefinite integral of f (that is, to find $\int f(x)\,\mathrm{d}x$ if x is chosen to be the independent variable).
- Same as that for definite integrals, in the notation $\int f(x)\,\mathrm{d}x$, the function f is called the *integrand*.

Integration of Constant (Function) Let k be a constant. Then we have

$$\int k\,\mathrm{d}x = kx + C, \qquad -\infty < x < \infty.$$

Explanation As usual, C is understood to be an arbitrary constant.

Proof The result follows from the Constant Multiple Rule for Differentiation and the Rule for Derivative of the Identity Function:

$$\begin{aligned}\frac{d}{dx}kx &= k \cdot \frac{d}{dx}x \\ &= k\end{aligned}$$

□

Example $\int 3\,dx = 3x + C$

Power Rule for Integration (positive integer version) Let n be a positive integer. Then we have

$$\int x^n\,dx = \frac{x^{n+1}}{n+1} + C, \qquad -\infty < x < \infty.$$

Proof The result follows from the Constant Multiple Rule and Power Rule (positive integer version) for Differentiation:

$$\begin{aligned}\frac{d}{dx}\frac{x^{n+1}}{n+1} &= \frac{1}{n+1} \cdot \frac{d}{dx}x^{n+1} \\ &= \frac{1}{n+1} \cdot (n+1)x^{n+1-1} \\ &= x^n\end{aligned}$$

□

Example $\int x^3\,dx = \frac{x^{3+1}}{3+1} + C = \frac{1}{4} \cdot x^4 + C$

Remark In the formula for Integration of Constant, putting $k = 1$, we get

$$\int 1\,dx = x + C, \qquad -\infty < x < \infty.$$

By considering the constant function 1 as the function x^0, the above result can be written as

$$\int x^0\,dx = \frac{x^{0+1}}{0+1} + C, \qquad -\infty < x < \infty.$$

Thus the Power Rule $\int x^n\,dx = \frac{x^{n+1}}{n+1} + C$ is also valid for the case where $n = 0$.

Power Rule for Integration (negative integer version) Let n be a negative integer different from -1. Then we have

$$\int x^n\,dx = \frac{x^{n+1}}{n+1} + C, \qquad x \neq 0.$$

Explanation The result means that on the intervals $(-\infty, 0)$ and $(0, \infty)$, the function $\frac{x^{n+1}}{n+1}$ is an antiderivative for the function x^n.

Proof The result follows from the Constant Multiple Rule and Power Rule (negative integer version) for Differentiation. □

Example $\int \frac{1}{x^5}\,dx = \int x^{-5}\,dx = \frac{x^{-5+1}}{-5+1} + C = \frac{-1}{4x^4} + C$

FAQ What is $\int \frac{1}{x}\,dx$?

Answer You can't apply the Power Rule if $n = -1$. Note that $\dfrac{x^{-1+1}}{-1+1}$ is meaningless. You will learn a formula in Chapter 8. □

Power Rule for Integration ($n + \frac{1}{2}$ version) Let n be an integer. Then we have

$$\int x^{n+\frac{1}{2}}\,\mathrm{d}x = \frac{x^{n+\frac{3}{2}}}{n+\frac{3}{2}} + C, \qquad x > 0.$$

Proof The result follows from the Constant Multiple Rule and Power Rule ($n + \frac{1}{2}$ version) for Differentiation. □

Remark The above result can be written as

$$\int x^r\,\mathrm{d}x = \frac{x^{r+1}}{r+1} + C, \qquad x > 0,$$

where $r = n + \dfrac{1}{2}$ and n is an integer. In fact, the formula is valid for all real numbers $r \neq -1$ (see Chapter 10).

Example $\displaystyle\int \frac{1}{\sqrt{x}}\,\mathrm{d}x = \int x^{-\frac{1}{2}}\,\mathrm{d}x = \frac{x^{-\frac{1}{2}+1}}{-\frac{1}{2}+1} + C = 2x^{\frac{1}{2}} + C$

Constant Multiple Rule for Integration Let k be a constant and let f be a function that is continuous on an open interval (a, b). Then we have

$$\int kf(x)\,\mathrm{d}x = k\int f(x)\,\mathrm{d}x, \qquad a < x < b.$$

Proof The result follows from the Constant Multiple Rule for Differentiation. □

Example Find $\displaystyle\int 2x^7\,\mathrm{d}x$.

Explanation The question is to find the family of functions that are antiderivatives for the integrand (on some open intervals). The answer should be given in the form "*a function of x +C*". Usually, for integration problems, there is no need to mention the underlying open intervals. For the given problem, the function $2x^7$ is continuous on $\mathbb{R}$ and so it has antiderivatives on $\mathbb{R}$.

Solution

$$\begin{aligned}
\int 2x^7\,\mathrm{d}x &= 2\int x^7\,\mathrm{d}x && \text{Constant Multiple Rule}\\
&= 2\left(\frac{x^{7+1}}{7+1} + C\right) && \text{Power Rule}\\
&= \frac{1}{4}x^8 + 2C
\end{aligned}$$

□

Remark From the answer, we see that the function $\frac{1}{4}x^8$ is an antiderivative for the function $2x^7$ (on $\mathbb{R}$). Therefore, we can also write

$$\int 2x^7\,\mathrm{d}x = \frac{1}{4}x^8 + C.$$

Although the answers $\frac{1}{4}x^8 + 2C$ and $\frac{1}{4}x^8 + C$ look different, they represent the same family of functions. In general, to do integration, we can use rules and formulas to get an antiderivative for the integrand and then add a constant of integration.

Sum Rule for Integration (Term by Term Integration) Let f and g be functions that are continuous on an open interval (a, b). Then we have

$$\int \big[f(x) + g(x)\big]\,\mathrm{d}x = \int f(x)\,\mathrm{d}x + \int g(x)\,\mathrm{d}x, \qquad a < x < b.$$

Proof The result follows from the Sum Rule for Differentiation. □

Example Find $\int (1 + x^3)\,\mathrm{d}x$.

Explanation We use rules and formulas for integration to obtain an antiderivative for the integrand and then add a constant of integration.

Solution

$$\begin{aligned} \int (1 + x^3)\,\mathrm{d}x &= \int 1\,\mathrm{d}x + \int x^3\,\mathrm{d}x && \text{Term by Term Integration} \\ &= x + \frac{x^4}{4} + C && \text{Power Rule} \end{aligned}$$

Remark Using the Sum Rule together with the Constant Multiple Rule, we obtain the following:

$$\int \big[f(x) - g(x)\big]\,\mathrm{d}x = \int f(x)\,\mathrm{d}x - \int g(x)\,\mathrm{d}x.$$

More generally, Term by Term Integration can be applied to sum and difference of finitely many terms.

Example Perform the following integration:

(1) $\displaystyle\int \left(x - 11 + \frac{3}{\sqrt{x}}\right)\mathrm{d}x$

(2) $\displaystyle\int (2x - 3)(x^2 + 1)\,\mathrm{d}x$

Explanation The question is to find the given indefinite integrals. The answers should be given in the form "*a function of x +C*".

Solution

(1)
$$\begin{aligned} \int \left(x - 11 + \frac{3}{\sqrt{x}}\right)\mathrm{d}x &= \int x\,\mathrm{d}x - \int 11\,\mathrm{d}x + \int 3x^{-\frac{1}{2}}\,\mathrm{d}x && \text{Term by Term Integration} \\ &= \frac{x^2}{2} - 11x + 3\int x^{-\frac{1}{2}}\,\mathrm{d}x && \text{Power Rule, Integration of Constant \& Constant Multiple Rule} \\ &= \frac{x^2}{2} - 11x + 3\cdot\frac{x^{\frac{1}{2}}}{\frac{1}{2}} + C && \text{Power Rule} \\ &= \frac{1}{2}x^2 - 11x + 6\sqrt{x} + C \end{aligned}$$

Remark In the second step, there is no need to add a constant of integration (because there is an indefinite integral in the third term). In the third step, we must add a constant of integration (otherwise, the expression represents a function but not a family of functions).

(2)
$$\begin{aligned}
\int (2x-3)(x^2+1)\,\mathrm{d}x &= \int (2x^3-3x^2+2x-3)\,\mathrm{d}x && \text{Rewrite the integrand}\\
&= \int 2x^3\,\mathrm{d}x - \int 3x^2\,\mathrm{d}x + \int 2x\,\mathrm{d}x - \int 3\,\mathrm{d}x && \text{Term by Term Integration}\\
&= 2\int x^3\,\mathrm{d}x - 3\int x^2\,\mathrm{d}x + 2\int x\,\mathrm{d}x - 3x && \text{Constant Multiple Rule \& Integration of Constant}\\
&= 2\cdot\frac{x^4}{4} - 3\cdot\frac{x^3}{3} + 2\cdot\frac{x^2}{2} - 3x + C && \text{Power Rule}\\
&= \frac{x^4}{2} - x^3 + x^2 - 3x + C
\end{aligned}$$

□

Caution $\int \left[f(x)\cdot g(x)\right]\mathrm{d}x \neq \int f(x)\,\mathrm{d}x \cdot \int g(x)\,\mathrm{d}x$

FAQ Do we have a rule for integration that corresponds to the product rule in differentiation?

Answer In integration, corresponding to the product rule, there is a technique called *integration by parts*. A brief introduction to this technique will be given in Chapter 10. □

To close this section, we give an example to illustrate the steps for finding definite integrals using rules for integration.

Example Evaluate the following definite integrals:

(1) $\displaystyle\int_{-1}^{2}(x^2-2x+3)\,dx$

(2) $\displaystyle\int_{0}^{1} x\,(x^2+1)\,dx$

Solution

(1)
$$\begin{aligned}
\int_{-1}^{2}(x^2-2x+3)\,dx &= \left[\frac{x^3}{3} - 2\cdot\frac{x^2}{2} + 3x\right]_{-1}^{2} && \text{Term by Term Integration, Power Rule, Constant Multiple Rule \& Fundamental Theorem of Calculus}\\
&= \left(\frac{8}{3} - 4 + 6\right) - \left(\frac{-1}{3} - 1 - 3\right)\\
&= 9
\end{aligned}$$

(2)
$$\begin{aligned}
\int_{0}^{1} x\,(x^2+1)\,dx &= \int_{0}^{1}(x^3+x^2)\,dx && \text{Rewrite the integrand}\\
&= \left[\frac{x^4}{4} + \frac{x^3}{3}\right]_{0}^{1} && \text{Term by Term Integration, Power Rule \& Fundamental Theorem of Calculus}\\
&= \frac{1}{4} + \frac{1}{3}\\
&= \frac{7}{12}
\end{aligned}$$

□

Exercise 6.3

1. Perform the following integration:

(a) $\int 2x^5\,\mathrm{d}x$ (b) $\int (3 - \frac{4}{\sqrt{x}})\,\mathrm{d}x$

(c) $\int (x^7 - 3x + 2)\,\mathrm{d}x$ (d) $\int (x^2 - \sqrt{x} + 3)\,\mathrm{d}x$

(e) $\int \frac{2}{3x\sqrt{x}}\,\mathrm{d}x$ (f) $\int (x^2 - 5x + 1)(2 - 3x)\,\mathrm{d}x$

(g) $\int (x^2 - 3)^2\,\mathrm{d}x$ (h) $\int \frac{x^2 + 1}{x^2}\,\mathrm{d}x$

2. Evaluate the following definite integrals:

(a) $\int_0^3 2x^3\,\mathrm{d}x$ (b) $\int_{-3}^3 2x^3\,\mathrm{d}x$

(c) $\int_{-1}^2 (1 - 5x^4)\,\mathrm{d}x$ (d) $\int_0^2 (x^4 - 3x^2 + 5)\,\mathrm{d}x$

(e) $\int_{-2}^2 (x^4 - 3x^2 + 5)\,\mathrm{d}x$ (f) $\int_1^4 (x^2 + \frac{1}{2\sqrt{x}})\,\mathrm{d}x$

(g) $\int_1^2 \frac{x^2 + 1}{\sqrt{x}}\,\mathrm{d}x$ (h) $\int_0^2 x(2 - 3x)^2\,\mathrm{d}x$

6.4 Application of Integration

Area under Graph of Function Let f be a function that is continuous on a closed and bounded interval $[a, b]$. Suppose that f is nonnegative on $[a, b]$. Then the area A of the region that lies below the graph of f and above the x-axis from $x = a$ to $x = b$ is given by

$$A = \int_a^b f(x)\,\mathrm{d}x.$$

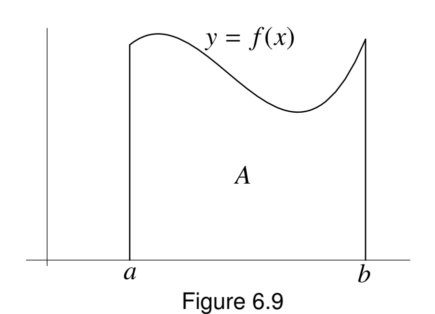


Figure 6.9

Example Find the area of the region that is bounded by the curve $y = \sqrt{x}$, the line $x = 1$ and the x-axis.

Explanation The curve, the vertical line and the x-axis divide the plane into six regions—five of them are unbounded (R_1, R_2, R_4, R_5 and R_6) and one of them is bounded (R_3, the required one).

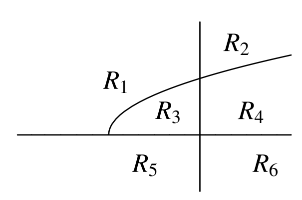


Figure 6.10a

Solution Let $f(x) = \sqrt{x}$. The region under consideration lies below the graph of f and above the x-axis from $x = 0$ to $x = 1$. The required area A is

$$\begin{aligned} A &= \int_0^1 f(x)\,\mathrm{d}x \\ &= \int_0^1 x^{\frac{1}{2}}\,\mathrm{d}x \\ &= \left[\frac{x^{\frac{3}{2}}}{\frac{3}{2}}\right]_0^1 \\ &= \frac{2}{3} \text{ (square units).} \end{aligned}$$

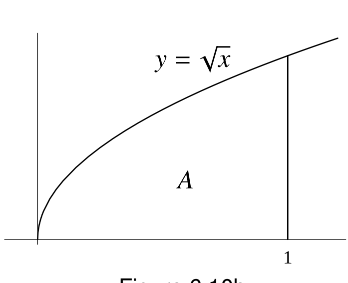


Figure 6.10b

□

Area between Graphs of Functions Let f and g be functions that are continuous on a closed and bounded interval $[a, b]$. Suppose that $f(x) \le g(x)$ for all $x \in [a, b]$ (this means that the graph of f lies below that of g). Then the area A of the region that is bounded by the graphs of f and g and the vertical lines $x = a$ and $x = b$ is given by

$$A = \int_a^b \left[g(x) - f(x)\right] \mathrm{d}x.$$

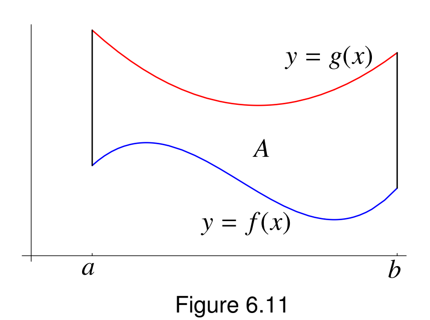


Figure 6.11

Proof For the case where f is nonnegative (hence the graphs of both f and g are above the x-axis), we have $A = A_g - A_f$, where A_g (respectively A_f) is the area of the region that lies below the graph of g (respectively the graph of f) and above the x-axis from $x = a$ to $x = b$ (see Figure 6.12a). Hence, using rules for definite integrals (Int1) and (Int2), we have

$$A = \int_a^b g(x)\,\mathrm{d}x - \int_a^b f(x)\,\mathrm{d}x = \int_a^b \left[g(x) - f(x)\right] \mathrm{d}x.$$

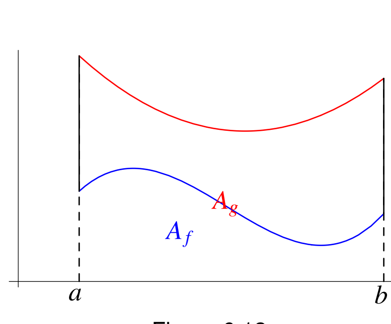


Figure 6.12a

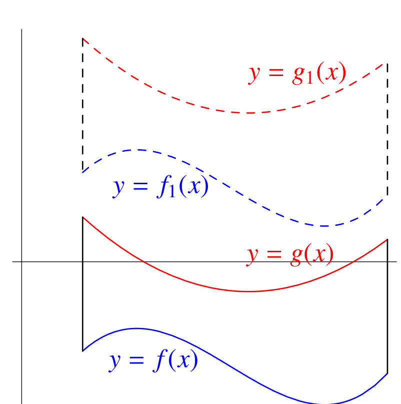


Figure 6.12b

For the general case, we can move the region upward suitably so that the graph of f is above the x-axis and then apply the result for the case where f is nonnegative (see Figure 6.12b). Indeed, since f is continuous on $[a, b]$, there exists a constant k such that $f(x) + k \ge 0$ for all $x \in [a, b]$. Let f_1 and g_1 be the functions from $[a, b]$ into $\mathbb{R}$ given by

$$f_1(x) = f(x) + k \quad \text{and} \quad g_1(x) = g(x) + k \qquad \text{for } a \le x \le b.$$

Since area is translation invariant, the required area A is equal to the area of the region that is bounded by the graphs of f_1 and g_1 and the vertical lines $x = a$ and $x = b$. Hence by what we obtain for the special case (since f_1 is nonnegative), we have

$$A = \int_a^b \left[g_1(x) - f_1(x)\right] \mathrm{d}x = \int_a^b \left[g(x) - f(x)\right] \mathrm{d}x.$$

□

Example Find the area of the region bounded by the parabola $y = x^2$ and the line $y = x + 2$.

Explanation The parabola and the line divide the plane into five regions—four of them are unbounded (R_1, R_2, R_3 and R_5) and one of them is bounded (R_4, the required one).

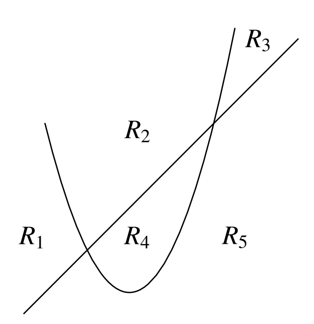


Figure 6.13a

Solution Solving for the x-coordinates of the intersection points of the parabola and the line:

$$\begin{aligned} x^2 &= x + 2 \\ x^2 - x - 2 &= 0 \\ (x-2)(x+1) &= 0, \end{aligned}$$

we get $x_1 = -1$ and $x_2 = 2$.

The region under consideration lies below the graph of $y = x + 2$, above that of $y = x^2$ (and between the vertical lines $x = -1$ and $x = 2$). The area A of the region is

$$\begin{aligned} A &= \int_{-1}^{2} \left[(x+2) - x^2\right] \mathrm{d}x \\ &= \left[\frac{x^2}{2} + 2x - \frac{x^3}{3}\right]_{-1}^{2} \\ &= \left(2 + 4 - \frac{8}{3}\right) - \left(\frac{1}{2} - 2 + \frac{1}{3}\right) \\ &= \frac{9}{2} \text{ (square units).} \end{aligned}$$

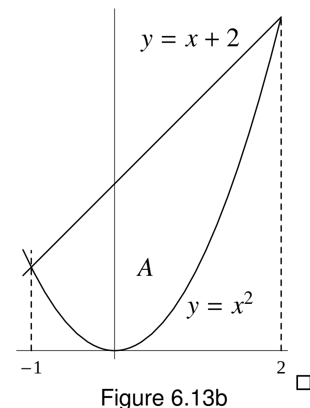


Figure 6.13b

□

Example Find the area of the combined region bounded by the curve $y = x^3 - 5x^2 + 6x$ and the x-axis.

Explanation The curve and the x-axis divide the plane into six regions—four of them are unbounded (R_1, R_3, R_4 and R_5) and two of them are bounded (R_2 and R_6). The two bounded regions intersect at one point and their union forms a combined region. The question is to find the area of $R_2 \cup R_6$.

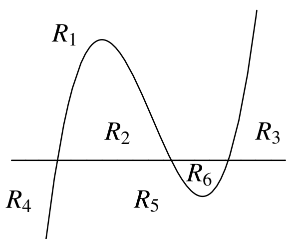


Figure 6.14a

Solution Solving for the x-coordinates of the intersection points of the curve and the x-axis:

$$\begin{aligned} x^3 - 5x^2 + 6x &= 0 \\ x(x^2 - 5x - 6) &= 0 \\ x(x-2)(x-3) &= 0, \end{aligned}$$

we get $x_1 = 0$, $x_2 = 2$ and $x_3 = 3$.

The required area A is $A = A_1 + A_2$ (see Figure 6.14b).

Note that for $0 \le x \le 2$, the curve is above the x-axis,
for $2 \le x \le 3$, the x-axis is above the curve.

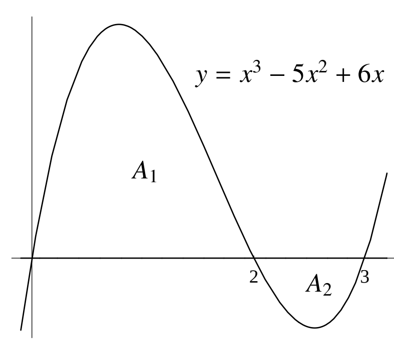


Figure 6.14b

Therefore, we have

$$\begin{aligned} A &= \int_0^2 [(x^3 - 5x^2 + 6x) - 0]\,\mathrm{d}x + \int_2^3 [0 - (x^3 - 5x^2 + 6x)]\,\mathrm{d}x \\ &= \left[\frac{x^4}{4} - \frac{5x^3}{3} + 3x^2\right]_0^2 - \left[\frac{x^4}{4} - \frac{5x^3}{3} + 3x^2\right]_2^3 \\ &= \left(\frac{8}{3} - 0\right) - \left(\frac{9}{4} - \frac{8}{3}\right) \\ &= \frac{37}{12}. \end{aligned}$$

□

Next we will give some examples that can be done using definite integrals as well as indefinite integrals. Before that, we give a result that is also known as the Fundamental Theorem of Calculus.

Fundamental Theorem of Calculus, Version 3 *Let f be a function such that f' is continuous on an open interval (a, b). Then for every $x_0 \in (a, b)$, we have*

$$f(x) = \int_{x_0}^{x} f'(t)\,\mathrm{d}t + f(x_0) \quad \textit{for all } x \in (a, b).$$

Proof Let g be the function from (a, b) into $\mathbb{R}$ defined by

$$g(x) = \int_{x_0}^{x} f'(t)\,\mathrm{d}t \qquad \text{for } a < x < b.$$

From the Fundamental Theorem of Calculus (Version 1), we see that g is an antiderivative for f' on (a, b). Since f is also an antiderivative of f' on (a, b), it follows from Theorem 6.3.1 that there exists a constant k such that

$$f(x) - g(x) = k \quad \text{for all } x \in (a, b).$$

Putting $x = x_0$, we get $f(x_0) - g(x_0) = k$ which yields $k = f(x_0)$ since $g(x_0) = \int_{x_0}^{x_0} f(t)\,\mathrm{d}t = 0$. Therefore we have

$$f(x) = g(x) + f(x_0) \quad \text{for all } x \in (a, b)$$

and the required result follows. □

Example Find an equation for the curve that passes through the point $(1, 0)$ and has slope function given by $x^3 - 2x + 1$.

Solution Let the curve be given by $y = f(x)$. Since the curve passes through the point $(1, 0)$, it follows that $f(1) = 0$. From the given slope function, we have $f'(x) = x^3 - 2x + 1$. Taking $x_0 = 1$ in the Fundamental Theorem of Calculus (Version 3), we have

$$\begin{aligned} f(x) &= \int_1^x f'(t)\,\mathrm{d}t + f(1) &&= \int_1^x (t^3 - 2t + 1)\,\mathrm{d}t + 0 \\ &= \left[\frac{t^4}{4} - t^2 + t\right]_1^x &&= \left(\frac{x^4}{4} - x^2 + x\right) - \left(\frac{1}{4} - 1 + 1\right) \\ &= \frac{x^4}{4} - x^2 + x - \frac{1}{4}. \end{aligned}$$

Therefore, an equation for the curve is: $y = \frac{x^4}{4} - x^2 + x - \frac{1}{4}$. □

Alternative solution The function f can also be found using indefinite integral:

$$f(x) = \int (x^3 - 2x + 1)\,\mathrm{d}x = \frac{x^4}{4} - x^2 + x + c$$

where the first inequality means that f is an antiderivative for $(x^3 - 2x + 1)$ and the second equality means that f is the function given by $\frac{x^4}{4} - x^2 + x + c$ and c is a specific constant (which is determined by f).

Putting $x = 1$, we get

$$0 = f(1) = \frac{1}{4} - 1 + 1 + c,$$

that is, $c = -\frac{1}{4}$. Therefore, we have $f(x) = \frac{x^4}{4} - x^2 + x - \frac{1}{4}$.

Example Find the cost function if the marginal cost is $3 + 40x - 5x^2$ and the fixed cost is 45.

Explanation Fixed cost is the cost when $x = 0$.

Solution

(*Method 1*) Let the cost function be C. By the Fundamental Theorem of Calculus (Version 3), we have

$$\begin{aligned} C(x) &= \int_0^x C'(t)\,\mathrm{d}t + C(0) \\ &= \int_0^x (3 + 40t - 5t^2)\,\mathrm{d}t + 45 \\ &= \left[3t + 20t^2 - \frac{5}{3}t^3\right]_0^x + 45 \\ &= 3x + 20x^2 - \frac{5}{3}x^3 + 45. \end{aligned}$$

(*Method 2*) Let the cost function be C. Then we have

$$C(x) = \int (3 + 40x - 5x^2)\,\mathrm{d}x = 3x + 20x^2 - \frac{5}{3}x^3 + c,$$

for some constant c. Putting $x = 0$, we get $45 = C(0) = c$ and so

$$C(x) = 3x + 20x^2 - \frac{5}{3}x^3 + 45.$$

□

The following result is a simple consequence of the Fundamental Theorem of Calculus (Version 3). It is known as the *Net Change Theorem* since $f(x_1) - f(x_0)$ is the net change of the values of f as x changes from x_0 to x_1.

Theorem 6.4.1 *Let f be a function such that f' is continuous on an open interval (a, b). Then for every pair of numbers x_0, x_1 in (a, b), we have*

$$\int_{x_0}^{x_1} f'(t)\,\mathrm{d}t = f(x_1) - f(x_0).$$

Example A particle moves along a line so that its velocity at time t is $v(t) = t^2 - t$ (measured in meters per second). Find the displacement of the particle during the time period $1 \le t \le 2$.

Explanation The question is to find $s(2) - s(1)$, where $s(t)$ is the position of the particle at time t. Note that the derivative of s is v.

Solution

(*Method 1*) By Theorem 6.4.1, the required displacement is

$$\begin{aligned} s(2) - s(1) &= \int_1^2 s'(t)\,\mathrm{d}t \\ &= \int_1^2 (t^2 - t)\,\mathrm{d}t \\ &= \left[\frac{t^3}{3} - \frac{t^2}{2}\right]_1^2 \\ &= \left(\frac{8}{3} - 2\right) - \left(\frac{1}{3} - \frac{1}{2}\right) \\ &= \frac{5}{6} \quad \text{(meter).} \end{aligned}$$

(*Method 2*) The displacement function of the particle is given by

$$s(t) = \int (t^2 - t)\,\mathrm{d}t = \frac{t^3}{3} - \frac{t^2}{2} + c,$$

where c is the initial position of the particle. The required displacement is

$$s(2) - s(1) = \left(\frac{8}{3} - 2 + c\right) - \left(\frac{1}{3} - \frac{1}{2} + c\right) = \frac{5}{6} \quad \text{(meter).}$$

□

Exercise 6.4

1. For each of the following, find the area of the region bounded by the given curve, the x-axis, and the given vertical line(s).
   (a) $y = x^3, \quad x = 3$
   (b) $y = x^2 - 4x, \quad x = 1, \quad x = 2$
   (c) $y = |x + 1| + 2, \quad x = -2, \quad x = 3$
   (d) $y = \sqrt{x + 3}, \quad x = 1$ *Hint: move the region appropriately.*
2. For each of the following, find the area of the (combined) region bounded by the given curves (or lines).
   (a) $y = \sqrt{x}$ and $y = x$
   (b) $y = x^2 - 4x - 8$ and $y = 2x - x^2$
   (c) $y = x$ and $y = x(x - 2)^2$
   (d) $y = x^2 - 4x + 4, \quad y = 10 - x^2$ and $y = 16$
3. Suppose f is a function such that $f'(x) = x^2 + 1$ and $f(1) = 2$. Find $f(x)$.
4. Suppose f is a function such that $f''(x) = (x + 1)(x - 2)$, $f(0) = 1$ and $f(1) = 0$. Find $f(x)$.
5. Water flows from the bottom of a storage tank at a rate of $r(t) = 150 - 5t$ liters per minute, where $0 \le t \le 30$. Find the amount of water that flows from the tank during the first 15 minutes.

Chapter 7

Trigonometric Functions

7.1 Angles

Idea of Definition An *angle* is formed by rotating a ray about its endpoint.

- The initial position of the ray is called the *initial side*.
- The endpoint of the ray is called the *vertex*.
- The final position is called the *terminal side*.

An angle is said to be in *standard position* if its vertex is at the origin and its initial side is along the positive x-axis.

Note An angle in standard position is uniquely determined by the direction and magnitude of rotation. So we can use numbers to represent angles.

- The direction of rotation may be counterclockwise or clockwise which will be considered to be positive or negative respectively.
- Magnitudes of rotation are traditionally measured in *degrees* where one revolution is defined to be 360 degrees, written $360°$.

Figures 7.1(a), (b) and (c) show three angles in standard position: Although the angles have the same terminal sides, their measures are different.

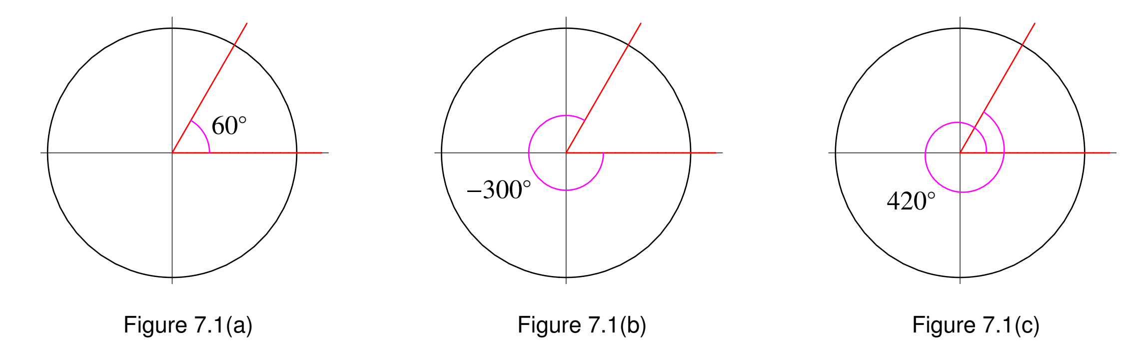


Figure 7.1(a) Figure 7.1(b) Figure 7.1(c)

Another unit for measuring angles is the radian. To define radian, we consider unit circles.

Terminology A circle with radius 1 is called a *unit circle*. The circle with radius 1 and center at the origin is called *the unit circle*.

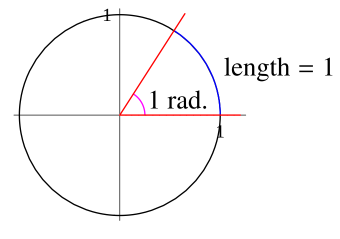


Figure 7.2

Definition The angle determined by an arc of length 1 along the circumference of a unit circle is said to be of measure *one radian*.

Since the circumference of a unit circle has length 2π, there are 2π radians in one revolution. Therefore, we have $360^\circ = 2\pi$ radians. The conversion between degrees and radians is given by

$$d^\circ = d \times \frac{\pi}{180} \text{ radians}$$

Thus, we have $90^\circ = \frac{\pi}{2}$ (radian) and $60^\circ = \frac{\pi}{3}$ (radian) for example.

Remark In calculus, it is more convenient to consider angles in radians and the unit *radian* is usually omitted.

Exercise 7.1

1. Convert the following degree measures to radians:
   (a) 270° (b) 210°
   (c) 315° (d) 750°
2. Convert the following radian measures to degrees:
   (a) $\frac{\pi}{6}$ (b) $\frac{3\pi}{4}$
   (c) $\frac{5\pi}{2}$ (d) 7π

7.2 Trigonometric Functions

Notation Consider an angle θ in standard position. Let P be the point of intersection of the terminal side and the unit circle. We define

$$\sin\theta = y\text{-coordinate of } P, \tag{7.2.1}$$

$$\cos\theta = x\text{-coordinate of } P. \tag{7.2.2}$$

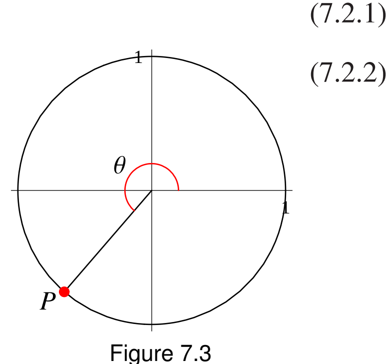


Figure 7.3

Remark Instead of considering the unit circle, we can also use circle of radius r (centered at the origin) and define $\cos\theta = \frac{a}{r}$ and $\sin\theta = \frac{b}{r}$, where a and b are the x- and y-coordinates of P respectively. It is easily seen (using similar triangles) that these ratios are independent of the choice of r.

Definition The rules given (7.2.1) and (7.2.2) define two functions from $\mathbb{R}$ into $\mathbb{R}$, called the *sine* and *cosine* functions respectively.

Using the sine and cosine functions, we define four more trigonometric functions, called the *tangent* (denoted by tan), *cotangent* (denoted by cot), *secant* (denoted by sec) and *cosecant* (denoted by csc) functions as

follows:

$$
\begin{aligned}
\tan x &= \frac{\sin x}{\cos x} && \text{provided that } \cos x \neq 0,\\
\cot x &= \frac{\cos x}{\sin x} && \text{provided that } \sin x \neq 0,\\
\sec x &= \frac{1}{\cos x} && \text{provided that } \cos x \neq 0,\\
\csc x &= \frac{1}{\sin x} && \text{provided that } \sin x \neq 0.
\end{aligned}
$$

Note Since $\sin x = 0$ if and only if $x = k\pi$ for some integer k and $\cos x = 0$ if and only if $x = \frac{k\pi}{2}$ for some odd integer k, it follows that

$$
\begin{aligned}
\operatorname{dom}(\tan) &= \operatorname{dom}(\sec) = \mathbb{R} \setminus \left\{ \pm\frac{\pi}{2}, \pm\frac{3\pi}{2}, \dots \right\},\\
\operatorname{dom}(\cot) &= \operatorname{dom}(\csc) = \mathbb{R} \setminus \left\{ \pm\pi, \pm 2\pi, \dots \right\}.
\end{aligned}
$$

Remark Below we will discuss some results for the sine, cosine and tangent functions. The secant function will only be used in an identity and a formula for differentiating the tangent function. The cotangent and cosecant functions will not be used in this course.

Properties

(1) The sine and cosine functions are periodic with period 2π, that is,

$$
\begin{aligned}
\sin(x + 2\pi) &= \sin x && \text{for all } x \in \mathbb{R},\\
\cos(x + 2\pi) &= \cos x && \text{for all } x \in \mathbb{R}.
\end{aligned}
$$

(2) The tangent function is periodic with period π, that is,

$$
\tan(x + \pi) = \tan x \qquad \text{for all } x \in \operatorname{dom}(\tan).
$$

(3) The sine function and the tangent function are *odd functions* and the cosine function is an *even function*, that is,

$$
\begin{aligned}
\sin(-x) &= -\sin x && \text{for all } x \in \mathbb{R},\\
\cos(-x) &= \cos x && \text{for all } x \in \mathbb{R},\\
\tan(-x) &= -\tan x && \text{for all } x \in \operatorname{dom}(\tan).
\end{aligned}
$$

(4) From (3), we see that the graphs of the sine function and tangent function are symmetric about the origin and the graph of the cosine function is symmetric about the y-axis. See Figures 7.4(a), (b) and (c).

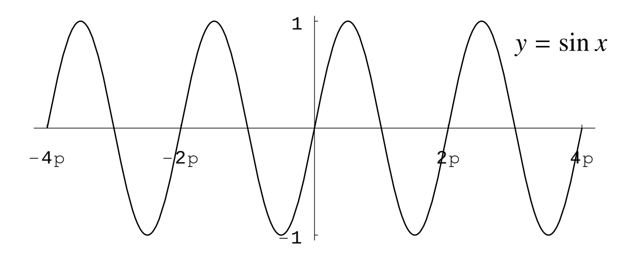


Figure 7.4(a)

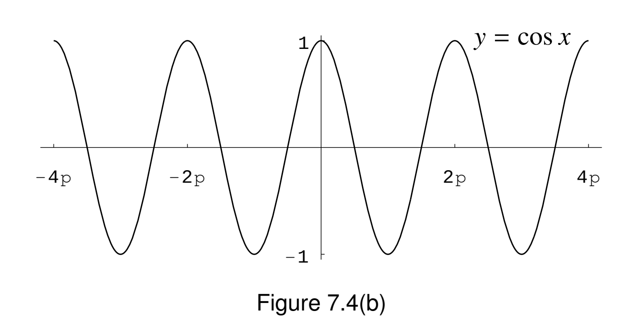


Figure 7.4(b)

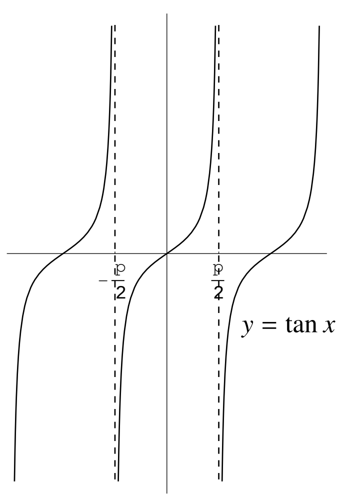


Figure 7.4(c)

Remark The graph of the cosine function can be obtain from that of the sine function by moving it $\frac{\pi}{2}$ units to the left. This is because

$$\cos x = \sin\left(x + \frac{\pi}{2}\right) \quad \text{for all } x \in \mathbb{R}.$$

CAST Rule The signs of sine, cosine and tangent in each quadrant can be memorized using the following rule:

S	A
T	C

where C stands for *cosine*, A for *all*, S for *sine* and T for *tangent*—for example, C in the 4th quadrant means that if x is an angle in the fourth quadrant, then $\cos x$ is positive and the other two values $\sin x$ and $\tan x$ are negative.

Sine, Cosine and Tangent of some Special Angles

$\sin 0 = 0$	$\cos 0 = 1$	$\tan 0 = 0$
$\sin \frac{\pi}{6} = \frac{1}{2}$	$\cos \frac{\pi}{6} = \frac{\sqrt{3}}{2}$	$\tan \frac{\pi}{6} = \frac{1}{\sqrt{3}}$
$\sin \frac{\pi}{4} = \frac{1}{\sqrt{2}}$	$\cos \frac{\pi}{4} = \frac{1}{\sqrt{2}}$	$\tan \frac{\pi}{4} = 1$
$\sin \frac{\pi}{3} = \frac{\sqrt{3}}{2}$	$\cos \frac{\pi}{3} = \frac{1}{2}$	$\tan \frac{\pi}{3} = \sqrt{3}$
$\sin \frac{\pi}{2} = 1$	$\cos \frac{\pi}{2} = 0$	$\tan \frac{\pi}{2}$ undefined

The values of the sine, cosine and tangent functions at the special angles can be obtained by drawing appropriate figures or triangles. For example, we can use Figures 7.5(a), (b) and (c) to find the values of the trigonometric functions for angles with size $\frac{\pi}{2}$, $\frac{\pi}{4}$ and $\frac{\pi}{6}$ respectively.

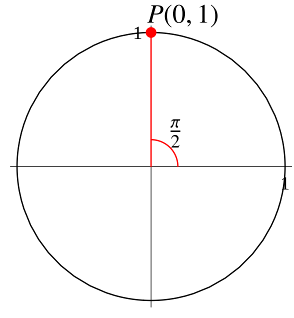


Figure 7.5(a)

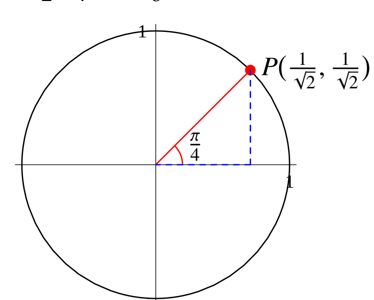


Figure 7.5(b)

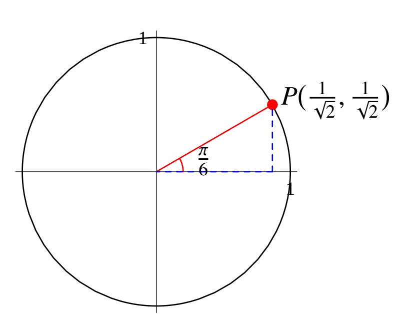


Figure 7.5(c)

An Important Identity

(Py) $\sin^2 x + \cos^2 x = 1$

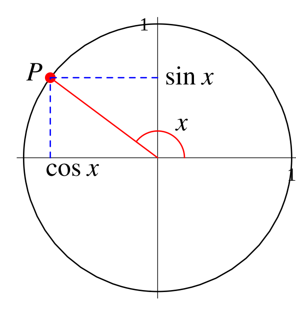


Figure 7.6

Explanation The result is called an identity because it is true for all $x \in \mathbb{R}$. Note that $\sin^2 x = (\sin x)^2$ etc.

Proof Let $P(a, b)$ be the point on the unit circle corresponding to angle x. By definition, we have

$$\sin x = b \quad \text{and} \quad \cos x = a.$$

The required result then follows since $a^2 + b^2 = 1$. □

Another Identity

(Py1) $1 + \tan^2 x = \sec^2 x, \quad x \in \mathbb{R} \setminus \{\pm\frac{\pi}{2}, \pm\frac{3\pi}{2}, \dots\}$

Proof

$$\begin{aligned} 1 + \tan^2 x &= 1 + \frac{\sin^2 x}{\cos^2 x} && \text{Definition of tan} \\ &= \frac{\cos^2 x + \sin^2 x}{\cos^2 x} \\ &= \frac{1}{\cos^2 x} && \text{Identity (Py)} \\ &= \left(\frac{1}{\cos x}\right)^2 \\ &= \sec^2 x && \text{Definition of sec} \end{aligned}$$

□

More Identities

$$\begin{aligned} \sin\left(\frac{\pi}{2} - x\right) &= \cos x & \cos\left(\frac{\pi}{2} - x\right) &= \sin x \\ \sin\left(\frac{\pi}{2} + x\right) &= \cos x & \cos\left(\frac{\pi}{2} + x\right) &= -\sin x \\ \sin(\pi - x) &= \sin x & \cos(\pi - x) &= -\cos x \\ \sin(\pi + x) &= -\sin x & \cos(\pi + x) &= -\cos x \\ \sin\left(\frac{3\pi}{2} - x\right) &= -\cos x & \cos\left(\frac{3\pi}{2} - x\right) &= -\sin x \\ \sin\left(\frac{3\pi}{2} + x\right) &= -\cos x & \cos\left(\frac{3\pi}{2} + x\right) &= \sin x \\ \sin(2\pi - x) &= -\sin x & \cos(2\pi - x) &= \cos x \end{aligned}$$

The above identities can be derived using appropriate figures. For example,

- from Figure 7.7(a), we get

$$\sin\left(\frac{\pi}{2} - x\right) = a = \cos x \quad \text{and} \quad \cos\left(\frac{\pi}{2} - x\right) = b = \sin x.$$

- from Figure 7.7(b), we get

$$\sin(\pi - x) = b = \sin x \quad \text{and} \quad \cos(\pi - x) = -a = -\cos x.$$

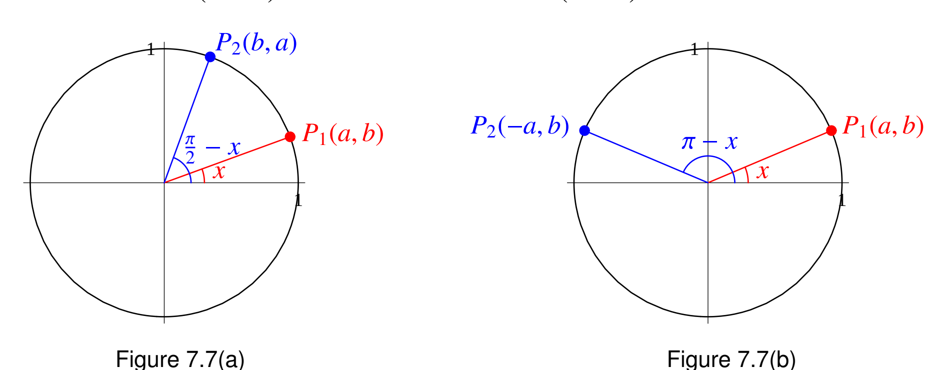


Figure 7.7(a) Figure 7.7(b)

Remark The above identities can be memorized in the following way:

$$f\left((\text{odd multiple of } \tfrac{\pi}{2}) \pm x\right) \quad \text{is} \quad \pm g(x)$$

$$f\left((\text{even multiple of } \tfrac{\pi}{2}) \pm x\right) \quad \text{is} \quad \pm f(x)$$

where $(f, g) = (\sin, \cos)$ or $(\cos, \sin)$ and the sign can be obtained by the CAST rule.

As illustrations, we describe how to obtaining the identities for

$$\sin(\pi - x), \qquad \cos(\pi - x), \qquad \sin\left(\frac{3\pi}{2} + x\right) \qquad \text{and} \qquad \cos\left(\frac{3\pi}{2} + x\right).$$

- Note that $\pi = 2 \cdot \frac{\pi}{2}$ is an even multiple of $\frac{\pi}{2}$. According to the second form (the trigonometric functions are unchanged),

  (1) $\sin(\pi - x)$ is either $\sin x$ or $-\sin x$

  (2) $\cos(\pi - x)$ is either $\cos x$ or $-\cos x$

  To determine the correct sign, we assume that x belongs to the 1st quadrant and so $(\pi - x)$ belongs to the 2nd quadrant. According to the CAST Rule, $\sin(\pi - x)$ is positive and $\cos(\pi - x)$ is negative (and $\sin x$ and $\cos x$ are positive). Thus we have

  (1c) $\sin(\pi - x) = \sin x$

  (2c) $\cos(\pi - x) = -\cos x$

- Note that $\frac{3\pi}{2}$ is an odd multiple of $\frac{\pi}{2}$. According to the first form (the two trigonometric functions are switched),

  (3) $\sin\left(\frac{3\pi}{2} + x\right)$ is either $\cos x$ or $-\cos x$

  (4) $\cos\left(\frac{3\pi}{2} + x\right)$ is either $\sin x$ or $-\sin x$

  To determine the correct sign, we assume that x belongs to the 1st quadrant and so $\left(\frac{3\pi}{2} + x\right)$ belongs to the 4th quadrant. According to the CAST Rule, $\sin\left(\frac{3\pi}{2} + x\right)$ is negative and $\cos\left(\frac{3\pi}{2} + x\right)$ is positive (and $\sin x$ and $\cos x$ are positive). Thus we have

  (3c) $\sin\left(\frac{3\pi}{2} + x\right) = -\cos x$

(4c) $\cos\left(\frac{3\pi}{2} + x\right) = \sin x$

Remark Since the values of the sine and cosine functions at $\frac{\pi}{2}$ or π or $\frac{3\pi}{2}$ can be found easily, the above identities can also be derived using the following results, called *compound angle formulas*.

Compound Angle Formulas Let A and B be real numbers. Then we have

$$\sin(A+B) = \sin A\cos B + \cos A \sin B \qquad \cos(A+B) = \cos A\cos B - \sin A\sin B$$
$$\sin(A-B) = \sin A\cos B - \cos A \sin B \qquad \cos(A-B) = \cos A\cos B + \sin A\sin B$$

Remark The formulas for $\sin(A-B)$ and $\cos(A-B)$ can be deduced from that for $\sin(A+B)$ and $\cos(A+B)$ respectively. This is because $\sin(-x) = -\sin x$ and $\cos(-x) = \cos x$ for all $x \in \mathbb{R}$. Moreover, since $\sin\left(\frac{\pi}{2} - x\right) = \cos x$ and $\cos\left(\frac{\pi}{2} - x\right) = \sin x$, the formula for $\sin(A+B)$ can be deduced from that for $\cos(A+B)$ and vice versa. However, the proof for either formula is very tedious and thus is omitted.

Continuity of sin and cos The sine and cosine functions are continuous on $\mathbb{R}$, that is, for every $a \in \mathbb{R}$, we have

$$\lim_{x\to a} \sin x = \sin a$$
$$\lim_{x\to a} \cos x = \cos a$$

Reason If x is close to a, then the point Q lying on the unit circle that corresponds to x is close to the point P that corresponds to a.

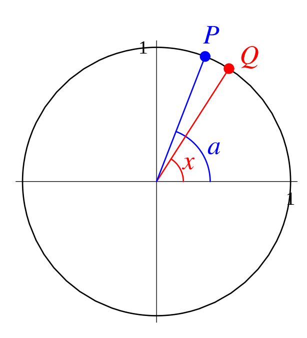


Figure 7.8

□

An Important Limit

(sin) $\displaystyle\lim_{x\to 0} \frac{\sin x}{x} = 1$

Proof First we consider right-side limit. Let x be small positive $(0 < x < \frac{\pi}{2})$. Consider the triangles $\triangle OAB$ and $\triangle OAC$ and the sector OAB shown in Figure 7.9. Note that

$$\text{area of } \triangle OAB = \frac{1}{2}\cdot 1\cdot 1\cdot \sin x = \frac{\sin x}{2}$$
$$\text{area of sector } OAB = \frac{1}{2}\cdot 1^2 \cdot x = \frac{x}{2}$$
$$\text{area of } \triangle OAC = \frac{1}{2}\cdot 1\cdot AC = \frac{\tan x}{2}$$

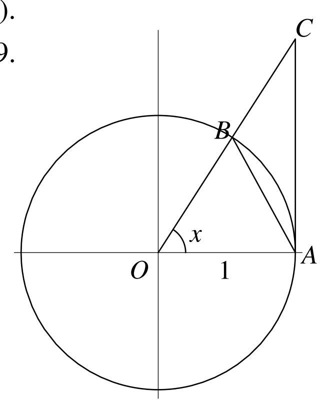


Figure 7.9

Since $\triangle OAB \subseteq$ sector $OAB \subseteq \triangle OAC$, it follows that

$$\frac{\sin x}{2} < \frac{x}{2} < \frac{\tan x}{2}.$$

Dividing each term by $\frac{\sin x}{2}$ (which is positive), we get

$$1 < \frac{x}{\sin x} < \frac{1}{\cos x},$$

which, by taking reciprocal, yields

$$1 > \frac{\sin x}{x} > \cos x. \tag{7.2.3}$$

By the continuity of the cosine function (at 0), we have

$$\lim_{x\to 0} \cos x = \cos 0 = 1. \tag{7.2.4}$$

Letting $x \to 0+$, by (7.2.3) and (7.2.4) together with the Sandwich Theorem (which is also valid for limits at a point and one-sided limits), we get

$$\lim_{x\to 0+} \frac{\sin x}{x} = 1.$$

Since $\frac{\sin x}{x}$ is an even function, that is, $\frac{\sin(-x)}{-x} = \frac{\sin x}{x}$ for all $x \neq 0$, it follows that $\lim_{x\to 0-} \frac{\sin x}{x} = 1$. Therefore, we have $\lim_{x\to 0} \frac{\sin x}{x} = 1$. □

Remark

- The result means that if x is small, then $\sin x$ is approximately equal to x.
- If x is in degrees, the result is different: $\lim_{x\to 0} \frac{\sin x^\circ}{x} = \frac{\pi}{180}$.

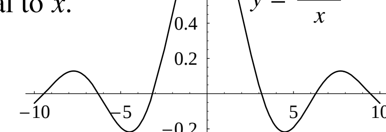


Figure 7.10

The following result will be used in deriving the formula for $\frac{\mathrm{d}}{\mathrm{d}x} \sin x$.

A Limit Result

(cos −1) $\lim_{h\to 0} \frac{\cos h - 1}{h} = 0$

Explanation To get the limit, we try to "cancel" the factor h in the denominator. However, the numerator is not a polynomial. Instead of making a factor h in the numerator, we try to make a factor $\sin h$ and apply (sin).

Proof

$$\begin{aligned}
\lim_{h\to 0} \frac{\cos h - 1}{h} &= \lim_{h\to 0} \frac{(\cos h - 1)(\cos h + 1)}{h(\cos h + 1)} && \text{Note: } \cos h + 1 \neq 0 \text{ if } h \approx 0 \\
&= \lim_{h\to 0} \frac{\cos^2 h - 1}{h(\cos h + 1)} \\
&= \lim_{h\to 0} \frac{-\sin^2 h}{h(\cos h + 1)} && \text{Identity (Py)} \\
&= \lim_{h\to 0} \left[(-\sin h) \cdot \frac{\sin h}{h} \cdot \frac{1}{\cos h + 1} \right] \\
&= \lim_{h\to 0} (-\sin h) \times \lim_{h\to 0} \frac{\sin h}{h} \times \lim_{h\to 0} \frac{1}{\cos h + 1} && \text{Limit Rule (La5)} \\
&= (-\sin 0) \cdot 1 \cdot \frac{1}{\cos 0 + 1} && \text{Continuity of sin \& cos and Limit Result (sin)} \\
&= 0.
\end{aligned}$$

□

Exercise 7.2

1. For each of the following, find its value without using calculators.

   (a) $\sin \frac{2\pi}{3}$ (b) $\cos \frac{2\pi}{3}$ (c) $\tan \frac{2\pi}{3}$

   (d) $\sin \frac{5\pi}{4}$ (e) $\cos \frac{5\pi}{4}$ (f) $\tan \frac{5\pi}{4}$

2. For each of the following limits, find its value.

   (a) $\lim_{x\to 0} \frac{\tan x}{x}$ (b) $\lim_{x\to 0} \frac{\sin 2x}{x}$

7.3 Differentiation of Trigonometric Functions

Derivative of sin The sine function is differentiable on $\mathbb{R}$ and its derivative is the cosine function, that is,

$$\frac{\mathrm{d}}{\mathrm{d}x} \sin x = \cos x, \qquad -\infty < x < \infty.$$

Proof

$$\begin{aligned}
\frac{\mathrm{d}}{\mathrm{d}x} \sin x &= \lim_{h\to 0} \frac{\sin(x+h) - \sin x}{h} && \text{Definition of Derivative} \\
&= \lim_{h\to 0} \frac{\sin x \cdot \cos h + \cos x \cdot \sin h - \sin x}{h} && \text{Compound Angle Formula} \\
&= \lim_{h\to 0} \frac{(\sin x \cdot \cos h - \sin x) + \cos x \cdot \sin h}{h} \\
&= \lim_{h\to 0} \frac{\sin x\,(\cos h - 1) + \cos x \cdot \sin h}{h} \\
&= \lim_{h\to 0} \left(\sin x \cdot \frac{\cos h - 1}{h}\right) + \lim_{h\to 0} \left(\cos x \cdot \frac{\sin h}{h}\right) && \text{Limit Rule (La4)} \\
&= \sin x \cdot \lim_{h\to 0} \frac{\cos h - 1}{h} + \cos x \cdot \lim_{h\to 0} \frac{\sin h}{h} && \text{Limit Rule (La5s)} \\
&= \sin x \cdot 0 + \cos x \cdot 1 && \text{Limit Results } (\cos -1) \text{ and } (\sin) \\
&= \cos x.
\end{aligned}$$

□

Derivative of cos The cosine function is differentiable on $\mathbb{R}$ and its derivative is the negative of the sine function, that is,

$$\frac{\mathrm{d}}{\mathrm{d}x} \cos x = -\sin x, \qquad -\infty < x < \infty.$$

Proof Similar to that for the derivative of the sine function, the result can be proved by definition, using the compound angle formula $\cos(x+h) = \cos x \cos h - \sin x \sin h$. □

Remark Note that $\cos x = \sin\left(\frac{\pi}{2} - x\right)$. The result can also be proved using the result for the derivative of the sine function. together with the chain rule which will be discussed in the Chapter 9.

Derivative of tan The tangent function is differentiable on its domain and its derivative is the square of the secant function, that is,

$$\frac{\mathrm{d}}{\mathrm{d}x} \tan x = \sec^2 x, \qquad x \neq \pm\frac{\pi}{2}, \pm\frac{3\pi}{2}, \ldots.$$

Proof For $x \in \operatorname{dom}(\tan) = \mathbb{R} \setminus \{\pm\frac{\pi}{2}, \pm\frac{3\pi}{2}, \ldots\}$, we have

$$\begin{aligned}
\frac{\mathrm{d}}{\mathrm{d}x}\tan x &= \frac{\mathrm{d}}{\mathrm{d}x}\left(\frac{\sin x}{\cos x}\right) && \text{Definition of tan} \\
&= \frac{\cos x \cdot \frac{\mathrm{d}}{\mathrm{d}x}\sin x - \sin x \cdot \frac{\mathrm{d}}{\mathrm{d}x}\cos x}{\cos^2 x} && \text{Quotient Rule} \\
&= \frac{\cos x \cdot \cos x - \sin x \cdot (-\sin x)}{\cos^2 x} && \text{Derivatives of sin and cos} \\
&= \frac{1}{\cos^2 x} && \text{Identity (Py)} \\
&= \left(\frac{1}{\cos x}\right)^2 && \\
&= (\sec x)^2 && \text{Definition of sec}
\end{aligned}$$

□

Example For each of the following y, find $\frac{\mathrm{d}y}{\mathrm{d}x}$.

(1) $y = 2\sin x - 7\cos x$

(2) $y = \tan x - x$

(3) $y = x\cos x$

(4) $y = \dfrac{\sin x}{x+1}$

Solution

(1)
$$\begin{aligned}
\frac{\mathrm{d}y}{\mathrm{d}x} &= \frac{\mathrm{d}}{\mathrm{d}x}(2\sin x - 7\cos x) && \text{Substitution} \\
&= 2 \cdot \frac{\mathrm{d}}{\mathrm{d}x}\sin x - 7 \cdot \frac{\mathrm{d}}{\mathrm{d}x}\cos x && \text{Term by Term Differentiation \& Constant Multiple Rule} \\
&= 2\cos x - 7(-\sin x) && \text{Derivative of sin and cos} \\
&= 2\cos x + 7\sin x
\end{aligned}$$

(2)
$$\begin{aligned}
\frac{\mathrm{d}y}{\mathrm{d}x} &= \frac{\mathrm{d}}{\mathrm{d}x}(\tan x - x) && \text{Substitution} \\
&= \frac{\mathrm{d}}{\mathrm{d}x}\tan x - \frac{\mathrm{d}}{\mathrm{d}x}x && \text{Term by Term Differentiation} \\
&= \sec^2 x - 1 && \text{Derivative of tan and Power Rule} \\
&= \tan^2 x && \text{Identity (Py1)}
\end{aligned}$$

(3)
$$\begin{aligned}
\frac{\mathrm{d}y}{\mathrm{d}x} &= \frac{\mathrm{d}}{\mathrm{d}x}(x^2\cos x) && \text{Substitution} \\
&= x^2 \cdot \frac{\mathrm{d}}{\mathrm{d}x}\cos x + \cos x \cdot \frac{\mathrm{d}}{\mathrm{d}x}x^2 && \text{Product Rule} \\
&= x^2 \cdot (-\sin x) + \cos x \cdot 2x && \text{Derivative of sin and Power Rule} \\
&= 2x\cos x - x^2\sin x
\end{aligned}$$

$$
\begin{aligned}
(4) \quad \frac{\mathrm{d}y}{\mathrm{d}x} &= \frac{\mathrm{d}}{\mathrm{d}x}\left(\frac{\sin x}{x+1}\right) && \text{Substitution} \\
&= \frac{(x+1)\cdot\frac{\mathrm{d}}{\mathrm{d}x}\sin x - \sin x\cdot\frac{\mathrm{d}}{\mathrm{d}x}(x+1)}{(x+1)^2} && \text{Quotient Rule} \\
&= \frac{(x+1)\cdot\cos x - \sin x\cdot(1+0)}{(x+1)^2} && \text{Derivative of sin and Derivative of Polynomial} \\
&= \frac{(x+1)\cos x - \sin x}{(x+1)^2}
\end{aligned}
$$

□

Exercise 7.3

1. For each of the following y, find $\frac{\mathrm{d}y}{\mathrm{d}x}$.
   (a) $y = 5\cos x - 2x$ (b) $y = 1 - 2\tan x$
   (c) $y = \sin x - x^2$ (d) $y = x^2 \sin x$
   (e) $y = \cos^2 x$ (f) $y = \frac{1}{\cos x}$
   (g) $y = \sin x \cdot \cos x$ (h) $y = \frac{\cos x}{x^3+1}$
   (i) $y = (x + \cos x)^2$ (j) $y = (\sin x + \cos x)^2$
2. Let $y = \sin^n x$.
   (a) Find $\frac{\mathrm{d}y}{\mathrm{d}x}$ for $n = 2, 3$ and 4.
   (b) Guess for formula for $\frac{\mathrm{d}y}{\mathrm{d}x}$ for general n (positive integer).
3. Let $y = \sin nx$ and let $z = \cos nx$.
   (a) Find $\frac{\mathrm{d}y}{\mathrm{d}x}$ and $\frac{\mathrm{d}z}{\mathrm{d}x}$ for $n = 2$. *Hint: use compound angle formulas.*
   (b) Find $\frac{\mathrm{d}y}{\mathrm{d}x}$ and $\frac{\mathrm{d}z}{\mathrm{d}x}$ for $n = 3$. *Hint: use compound angle formulas and the results in (a).*
   (c) Guess for formula for $\frac{\mathrm{d}y}{\mathrm{d}x}$ and $\frac{\mathrm{d}z}{\mathrm{d}x}$ for general n (positive integer).
4. Let $f(x) = \sin(ax + b)$ and let $g(x) = \cos(ax + b)$ where a and b are constants.
   (a) Use definition to find $f'(x)$ and $g'(x)$.
   (b) Use the results in (a) to find $f''(x)$ and $g''(x)$
   (c) Guess for formula for $f^{(n)}(x)$ and $g^{(n)}(x)$ for general n (positive integer).

Chapter 8

Exponential and Logarithmic Functions

8.1 Exponential Functions

Definition Let $0 < b \neq 1$. We define $\exp_b$ to be the function from $\mathbb{R}$ into $\mathbb{R}$ given by

$$\exp_b(x) = b^x, \qquad x \in \mathbb{R}.$$

The function $\exp_b$ is called the *exponential function* with base b.

Remark

- If x is a rational number, such as $x = \frac{4}{3}$, then $b^x = b^{\frac{4}{3}} = \sqrt[3]{b^4}$.
- If x is an irrational number such as $x = \sqrt{2}$, to define b^x we use approximations: more precisely we use limits.

  Example To assign a value to $3^{\sqrt{2}}$:

  Note that

  $$\sqrt{2} = 1.414213562373095\cdots$$

  We may use $3^{1.4} = 3^{\frac{14}{10}}$ to give an approximate value for $3^{\sqrt{2}}$. For better approximations, we may use $3^{1.41}$, $3^{1.414}$ and so on. Denote

  $$a_1 = 1.4, \quad a_2 = 1.41, \quad a_3 = 1.414, \quad a_4 = 1.4142, \quad a_5 = 1.41421, \quad \ldots$$

  It can be shown that the sequence

  $$3^{a_1},\ 3^{a_2},\ 3^{a_3},\ 3^{a_4},\ \ldots$$

  is convergent and $3^{\sqrt{2}}$ is defined to be the limit of the sequence.

n	3^{a_n}
1	4.655536722
2	4.706965002
3	4.727695035
4	4.728733930
5	4.728785881
6	4.728801466
7	4.728804064
8	4.728804376
9	4.728804386
⋮	

FAQ Instead of the above sequence (a_n), can we take other sequences (b_n) of rational numbers converging to $\sqrt{2}$ and use $\lim_{n\to\infty} 3^{b_n}$ to define $3^{\sqrt{2}}$?

Answer You can take any sequence (b_n) converging to $\sqrt{2}$. It can be shown (but difficult!) that $\lim_{n\to\infty} 3^{b_n}$ always exists and is independent of the choice of (b_n). □

FAQ In the definition of exponential functions, why do we exclude $b = 1$?

Answer When $b = 1$, the function $1^x = 1$ is trivial: a constant function. It does not enjoy the *injective* property possessed by $\exp_b$ where $b \neq 1$. When we define logarithmic functions, we need exponential functions be injective.

We need $b > 0$ because we want $\exp_b(x) = b^x$ to be defined for all real numbers x. If b is zero, b^{-1} is undefined; if b is negative, $b^{\frac{1}{2}}$ is undefined. □

Rules for Exponent Let a and b be positive real numbers different from 1. Then for every $x \in \mathbb{R}$ and every $y \in \mathbb{R}$, we have

(1) $a^x a^y = a^{x+y}$ (2) $\dfrac{a^x}{a^y} = a^{x-y}$

(3) $(a^x)^y = a^{xy}$ (4) $(ab)^x = a^x b^x$

(5) $\left(\dfrac{a}{b}\right)^x = \dfrac{a^x}{b^x}$ (6) $a^1 = a$

(7) $a^0 = 1$ (8) $a^{-x} = \dfrac{1}{a^x}$

Proof The results follow from the corresponding rules for that with rational exponents (see an FAQ on page 36).

Continuity of Exponential Functions By definition, the domain of every exponential function is $\mathbb{R}$. It can be shown that exponential functions are continuous on $\mathbb{R}$, that is, if x is close to x_0, then b^x is close to b^{x_0}.

Range of Exponential Functions Since $\exp_b(x) = b^x$ is always positive, it follows that the ranges of the exponential functions are contained in $(0, \infty)$. In fact, we have

$$\text{range}(\exp_b) = (0, \infty).$$

Proof For $b > 1$, since $\lim_{x\to\infty} b^x = \infty$ and $\lim_{x\to-\infty} b^x = 0$, it follows that the exponential function $\exp_b$ can attain arbitrarily large values as well as arbitrarily small positive values. Hence by the Intermediate Value Theorem, it can attain any positive value. Therefore, the range of $\exp_b$ is $(0, \infty)$.

For $0 < b < 1$, the range of $\exp_b$ is also $(0, \infty)$. This is because $\exp_b(x) = \exp_{\frac{1}{b}}(-x)$ by Rule for Exponents (8). □

Graph of Exponential Functions In general, the graph of an exponential function has one of two general shapes depending on the value of the base b.

Remark

- The graph of $y = b^x$ and that of $y = \left(\frac{1}{b}\right)^x$ are symmetric about the y-axis. This is because $b^{-x} = \left(\frac{1}{b}\right)^x$.
- As x increases, the graph goes up if $b > 1$ and goes down if $0 < 1 < b$.

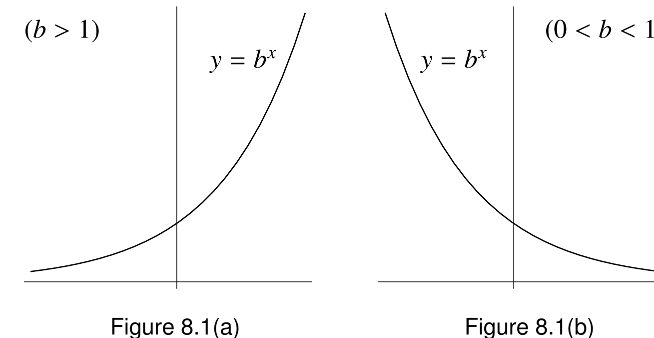


Figure 8.1(a) Figure 8.1(b)

Consider the expression s^t.

- If we let $s = b$ (> 0) be a fixed positive real number and let $t = x$ varies in $\mathbb{R}$, then we get a function $x \mapsto b^x$, which is the exponential function $\exp_b$ if $b \neq 1$ or the constant function 1 if $b = 1$.
- If we let $t = r$ be a fixed real number and let $s = x$ varies in $(0, \infty)$, we get a power function $x \mapsto x^r$.

Continuity of Power Functions It can be shown that for every real number r, the power function x^r is continuous on $(0, \infty)$.

Recall that a function f is said to be *injective* if the following condition is satisfied:

$(*)$ If $x_1, x_2 \in \operatorname{dom}(f)$ and $x_1 \neq x_2$, then $f(x_1) \neq f(x_2)$.

Condition $(*)$ is equivalent to the following condition:

$(**)$ If $x_1, x_2 \in \operatorname{dom}(f)$ and $f(x_1) = f(x_2)$, then $x_1 = x_2$.

If $\operatorname{dom}(f)$ is a subset of $\mathbb{R}$ and the codomain of f is $\mathbb{R}$, then f is injective means that the graph of f intersects every horizontal line in at most one point.

Injectivity of $\exp_b$ The exponential functions $\exp_b$ $(0 < b \neq 1)$ are injective.

Proof Let $s, t \in \mathbb{R}$ and $s \neq t$. Without loss of generality, we may assume that $s < t$. Then we have $b^s < b^t$ if $b > 1$ and $b^s > b^t$ if $0 < b < 1$. In any case, we have $\exp_b(s) \neq \exp_b(t)$. Thus, the exponential functions $\exp_b$ $(0 < b \neq 1)$ are injective. □

Remark The graph of $y = b^x$ intersects the horizontal line $y = c$ in exactly one point if $c > 0$ and in no point if $c \leq 0$.

Exponential Equations To solve simple equations involving exponentials, we use the fact that exponential functions are injective.

Example For each of the following equations, find its solution set.

(1) $3^{2x-1} = \dfrac{1}{3^{5-x}}$

(2) $8^{x^2} = 4^{x+4}$

Explanation To use the injective property of the exponential functions, we have to express both sides of the equation in the form $b^{\text{something}}$. For (1), we can take $b = 3$ and for (2), we can take $b = 2$.

Solution

(1)
$$
\begin{aligned}
3^{2x-1} &= \frac{1}{3^{5-x}} & \\
3^{2x-1} &= 3^{-(5-x)} & \text{Rewrite right-side} \\
3^{2x-1} &= 3^{x-5} & \\
2x - 1 &= x - 5 & \text{Injectivity of } \exp_3 \\
x &= -4 &
\end{aligned}
$$

The solution set is $\{-4\}$.

(2)
$$\begin{aligned}
8^{x^2} &= 4^{x+4} & \\
(2^3)^{x^2} &= (2^2)^{x+4} & \text{Rewrite both sides} \\
2^{3x^2} &= 2^{2(x+4)} & \\
3x^2 &= 2(x+4) & \text{Injectivity of } \exp_2 \\
3x^2 - 2x - 8 &= 0 & \\
(x-2)(3x+4) &= 0 &
\end{aligned}$$

The solution set is $\{2, -\frac{4}{3}\}$.

□

The number e It can be shown that $\lim_{x\to\infty}\left(1+\frac{1}{x}\right)^x$ exists. The following table and figure illustrate this fact (the proof of the fact is beyond the scope of this course). This limit will be denoted by e, that is,

$$\mathrm{e} = \lim_{x\to\infty}\left(1+\frac{1}{x}\right)^x.$$

x	$(1+1/x)^x$
10	2.59374
100	2.70481
1000	2.71692
10000	2.71815
20000	2.71821
30000	2.71824
40000	2.71825

Figure 8.2

Remark It can be shown that e is an irrational number (the proof is not easy!). The following gives the value of e correct to 50 decimal places:

$$2.71828182845904523536028747135266249775724709369996\ldots$$

Notation and Terminology We write exp (omitting the base) to denote the exponential function with base e. When we say *the exponential function*, we mean the function exp.

Remark As usual, we also write e^x to denote the exponential function with base e.

Below we discuss two situations in which the number e appears.

Interests Compounded Continuously When money is invested at a given annual rate, the interest earned depends on how frequently interest is compounded.

Consider a principal of P dollars invested for t years at an annual rate of r. If interest is compounded k times a year, then the rate per conversion period is $\frac{r}{k}$ and there are kt periods. The compounded amount $A(t)$ at the end of t years is given by

$$A(t) = P\left(1+\frac{r}{k}\right)^{kt}.$$

If $k \to \infty$, the number of conversion periods increases indefinitely and the length of each period approaches 0. In this case, we say that interest is *compounded continuously*. The compounded amount $A_c(t)$ at the end of t years is

$$\begin{aligned} A_c(t) &= \lim_{k\to\infty} P\left(1+\frac{r}{k}\right)^{kt} \\ &= P\lim_{k\to\infty}\left(1+\frac{r}{k}\right)^{\frac{k}{r}\cdot rt} && \text{Limit Rule (L5s) and rewrite exponent} \\ &= P\left[\lim_{k\to\infty}\left(1+\frac{r}{k}\right)^{\frac{k}{r}}\right]^{rt} && \text{Continuity of Power Function} \\ &= P\left[\lim_{x\to\infty}\left(1+\frac{1}{x}\right)^{x}\right]^{rt} && \text{Put } x=\frac{k}{r} \end{aligned}$$

Note that the limit inside the brackets is the number e. Therefore we have the following formula:

$$A_c(t) = P\mathrm{e}^{rt}.$$

Radioactive Decay Suppose that the initial amount (at $t = 0$) of a radioactive substance is A_0. Then the amount $A(t)$ of the substance at time t is given by

$$A(t) = A_0\mathrm{e}^{-\lambda t},$$

where λ is a (positive) constant, called the *decay constant* of the substance.

Remark For example, the decay constant of carbon 14 is about 0.00012.

(1) To find the amount at a certain time t, we can just plug in the value of t. If we want to find the time so that the amount is reduced to A_1, we need to solve the following equation for t

$$A_1 = A_0\mathrm{e}^{-\lambda t}.$$

This will be discussed in the next section.

(2) The *half-life* of a radioactive element is the length of time required for a given quantity of the element to decay to one-half of its original mass. For example, the half-life of carbon 14 is about 5730 years.

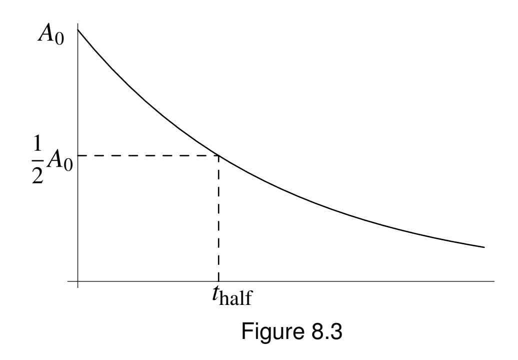


Figure 8.3

Exercise 8.1

1. For each of the following, sketch the graphs of the given equations on the same coordinate plane.
   (a) $y = 2^x$, $y = 2.5^x$, $y = 3^x$
   (b) $y = 2^x$, $y = (\frac{1}{2})^x$
   (c) $y = 2^x$, $y = 2^x + 1$, $y = 2^x - 2$
   (d) $y = 2^x$, $y = 2^{x+1}$, $y = 2^{x-2}$
2. For each of the function f, find its domain and range.
   (a) $f(x) = 3^x + 1$ (b) $f(x) = \frac{1}{3^x + 1}$
   (c) $f(x) = \frac{1}{3^x - 1}$

3. For each of the following equations, find its solution set.
   (a) $2^{2x} = 2^{x^2-3}$ (b) $e^{x+2} = 1$
   (c) $e^{x+2} = 0$ (d) $x^2e^x = 2xe^x$

8.2 Logarithmic Functions

For each $b > 0$ with $b \neq 1$, the exponential function $\exp_b$ is injective and its range is $(0, \infty)$. Thus it has an inverse whose domain is $(0, \infty)$.

Definition Let $0 < b \neq 1$. We define $\log_b$ to be the inverse of the exponential function $\exp_b$. The function $\log_b$ is called the *logarithmic function* with base b.

Remark $\log_b$ is the function from $(0, \infty)$ into $\mathbb{R}$ such that the following two conditions are satisfied:

(Log1) $\log_b(b^x) = x$ for all $x \in \mathbb{R}$;

(Log2) $b^{\log_b y} = y$ for all $y \in (0, \infty)$.

By the definition of inverse, we have the following:

$$x = \log_b y \qquad \text{means} \qquad y = b^x. \tag{8.2.1}$$

Example For each of the following, convert it to an equivalent logarithmic form.

(1) $5^2 = 25$

(2) $10^0 = 1$

Solution Using (8.2.1), we obtain

(1) $\log_5 25 = 2$

(2) $\log_{10} 1 = 0$

Example For each of the following, convert it to an equivalent exponential form.

(1) $\log_{10} 1000 = 3$

(2) $\log_2 \frac{1}{16} = -4$

Solution Using (8.2.1), we obtain

(1) $10^3 = 1000$

(2) $2^{-4} = \frac{1}{16}$

Continuity of $\log_b$ It can be shown that logarithmic functions are continuous on $(0, \infty)$.

Graphs of Logarithmic Functions For every $b > 0$ with $b \neq 1$, since the logarithmic function $\log_b$ is the inverse of the exponential function $\exp_b$, it follows that the graph of $\log_b$ and that of $\exp_b$ are symmetric about the line $x = y$.

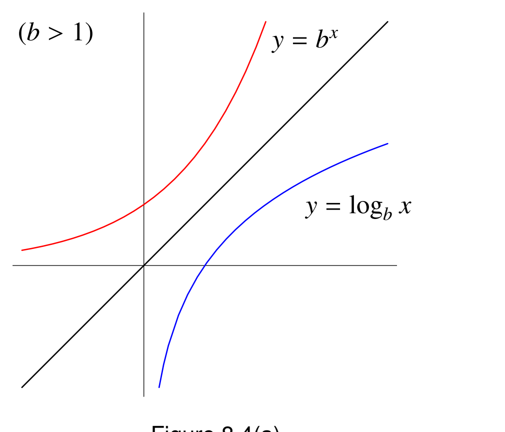


Figure 8.4(a)

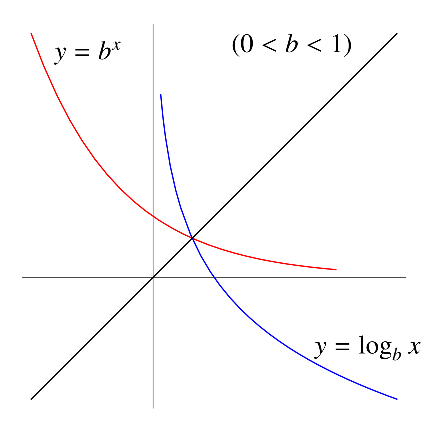


Figure 8.4(b)

Terminology and Notation Logarithms with the base 10 are called *common logarithms.* They were frequently used for computational purposes before the calculator age. The subscript 10 is usually omitted:

$$\log x \quad \text{means} \quad \log_{10} x.$$

In calculus, logarithms with base e, called *natural logarithms*, are more important. We use the notation "ln" for such logarithms:

$$\ln x \quad \text{means} \quad \log_{\mathrm{e}} x.$$

Caution Many students write ln as In (which, of course, is not correct). To avoid this, some authors write ℓn for natural logarithm. Note that the letter "l" comes from the word "logarithm" and the letter "n" comes from "natural".

Remark In many advanced books, common logarithm is never used and the symbol "log" stands for natural logarithm.

Properties of Logarithms

(1) $\log_b 1 = 0$

(2) $\log_b b = 1$

(3) $\log_b(mn) = \log_b m + \log_b n$ (logarithm of product is sum of logarithms)

(4) $\log_b \dfrac{1}{m} = -\log_b m$

(5) $\log_b \dfrac{m}{n} = \log_b m - \log_b n$ (logarithm of quotient is difference of logarithms)

(6) $\log_b m^r = r \log_b m$

(7) $\log_b b^r = r$

(8) $b^{\log_b m} = m$

(9) $\log_b m = \dfrac{\log_a m}{\log_a b}$ (change of base formula)

where m, n, a, b are positive real numbers with a and b different from 1 and r can be a real number.

Proof

(1) Since $b^0 = 1$, it follows from (8.2.1) that $\log_b 1 = 0$.

(2) Since $b^1 = b$, it follows from (8.2.1) that $\log_b b = 1$.

(3) Denote $x = \log_b m$ and $y = \log_b n$. By (8.2.1), we have $b^x = m$, $b^y = n$ which implies that

$$\begin{aligned} mn &= b^x b^y \\ &= b^{x+y} \qquad \text{Rule of Exponent (1).} \end{aligned}$$

Again by (8.2.1), we have $\log_b(mn) = x + y$ from which we get the required equality.

(4) Note that
$$\begin{aligned} \log_b m + \log_b \frac{1}{m} &= \log_b \left(m \cdot \frac{1}{m}\right) \qquad \text{Property (3)} \\ &= \log_b 1 \\ &= 0 \qquad \text{Property (1).} \end{aligned}$$
Thus the required equality follows.

(5) Apply Properties (3) and (4).

(6) Denote $y = \log_b m$. By (8.2.1), we have $b^y = m$ which implies that

$$\begin{aligned} m^r &= (b^y)^r \\ &= b^{ry} \qquad \text{Rule of Exponent (3).} \end{aligned}$$

Again by (8.2.1), we have $\log_b m^r = ry$ from which we get the required equality.

(7) The result follows from the definition of logarithmic functions. See (Log1). Alternatively, we may apply Properties (6) and (2).

(8) The result follows from the definition of logarithmic functions. See (Log2).

(9) Denote $x = \log_b m$. By (8.2.1), we have $b^x = m$. Take the logarithm to the base a of both sides, we get

$$\log_a b^x = \log_a m.$$

By Property (6), we have $x \log_a b = \log_a m$, which implies that $x = \frac{\log_a m}{\log_a b}$. The required equality then follows.

□

Caution In general,
$$\begin{aligned} &\log_b(m + n) \neq \log_b m \times \log_b n \qquad \text{See Property (3) for the correct form.} \\ &\log_b(m - n) \neq \log_b m \div \log_b n \qquad \text{See Property (5) for the correct form.} \end{aligned}$$

Example For each of the following, find its values (without using calculators).

(1) $\log_6 54 - \log_6 9$

(2) $e^{4 \ln 3 - 3 \ln 4}$

Solution

(1)
$$\begin{aligned} \log_6 54 - \log_6 9 &= \log_6 \frac{54}{9} \qquad \text{Property (5)} \\ &= \log_6 6 \\ &= 1 \qquad \text{Property (2)} \end{aligned}$$

(2)
$$\begin{aligned} e^{4\ln 3 - 3\ln 4} &= e^{\ln 3^4 - \ln 4^3} && \text{Property (6)} \\ &= e^{\ln 81 - \ln 64} && \text{Property (5)} \\ &= e^{\ln \frac{81}{64}} \\ &= \frac{81}{64} && \text{Property (8)} \end{aligned}$$

□

Exponential and Logarithmic Equations To solve simple equations involving logarithms and exponentials, we may use the following methods:

- Apply (8.2.1) to change logarithmic form to its equivalent exponential form or vice versa.
- Use the fact that logarithmic functions and exponential functions are injective (examples of simple equations involving are given in the last section).

  Remark The inverse function of every injective function is injective. In particular, logarithmic functions are injective.

Example For each of the following equations, find its solution set.

(1) $\log_2 x = -3$

(2) $\ln(2x + 1) = 4$

(3) $\log_x 49 = 2$

(4) $e^{3x} = 14$

Solution

(1)
$$\begin{aligned} \log_2 x &= -3 \\ 2^{-3} &= x && \text{by (8.2.1)} \\ \frac{1}{8} &= x \end{aligned}$$

The solution set is $\{\frac{1}{8}\}$.

(2)
$$\begin{aligned} \ln(2x + 1) &= 4 \\ e^4 &= 2x + 1 && \text{by (8.2.1)} \\ e^4 - 1 &= 2x \end{aligned}$$

The solution set is $\{\frac{e^4 - 1}{2}\}$.

(3)
$$\begin{aligned} \log_x 49 &= 2 \\ x^2 &= 49 \quad \text{and} \quad x > 0 && \text{by (8.2.1) and condition of base} \\ x &= \pm 7 \quad \text{and} \quad x > 0 \end{aligned}$$

The solution set is $\{7\}$.

(4)
$$\begin{aligned} e^{3x} &= 14 \\ \ln 14 &= 3x \quad \text{and} \quad x > 0 && \text{by (8.2.1)} \end{aligned}$$

The solution set is $\{\frac{\ln 14}{3}\}$.

□

Example Solve the equation $\log_2 x = 5 - \log_2(x+4)$.

Explanation To apply (8.2.1), we have to rewrite the equation in the form $\log_b m = n$. Note that in the expressions $\log_2 x$ and $\log_2(x+4)$, it is assumed that $x > 0$ and $x + 4 > 0$ respectively.

Alternatively, we may use the fact that logarithmic functions are injective. In order to apply this, we have to express both sides of the equation in the form $\log_b$(something).

Solution 1

$$\begin{aligned}
\log_2 x &= 5 - \log_2(x+4) \\
\log_2 x + \log_2(x+4) &= 5 \\
\log_2 x(x+4) &= 5 \quad \text{and} \quad x > 0 \quad \text{and} \quad x+4 > 0 && \text{Property (3) \& Domain of } \log_2 \\
x(x+4) &= 2^5 \quad \text{and} \quad x > 0 && \text{by (8.2.1)} \\
x^2 + 4x - 32 &= 0 \quad \text{and} \quad x > 0 \\
(x+8)(x-4) &= 0 \quad \text{and} \quad x > 0
\end{aligned}$$

The solution set is {4}. □

Solution 2

$$\begin{aligned}
\log_2 x &= \log_2 2^5 - \log_2(x+4) && \text{Property (6)} \\
\log_2 x &= \log_2 \frac{2^5}{x+4} && \text{Property (5)} \\
x &= \frac{2^5}{x+4} \quad \text{and} \quad x > 0 && \text{Injectivity of } \log_2 \text{ \& Domain of } \log_2 \\
x(x+4) &= 32 \quad \text{and} \quad x > 0 \\
x^2 + 4x - 32 &= 0 \quad \text{and} \quad x > 0 \\
(x+8)(x-4) &= 0 \quad \text{and} \quad x > 0
\end{aligned}$$

The solution set is {4}. □

Half-Life To find the half-life of a radioactive element, we have to solve

$$\frac{1}{2}A_0 = A_0 e^{-\lambda t}.$$

Taking ln on both sides, or equivalently, using (8.2.1), we get

$$\begin{aligned}
\ln \frac{1}{2} &= -\lambda t \\
-\ln 2 &= -\lambda t && \text{Property (4)} \\
t &= \frac{\ln 2}{\lambda}.
\end{aligned}$$

This gives the relation between the half-life t_{half} and the decay constant λ.

Exercise 8.2

1. For each of the following, convert it to an equivalent logarithmic form.
   (a) $9^2 = 81$ (b) $2 = \sqrt{4}$ (c) $\frac{1}{2} = 2^{-1}$
2. For each of the following, convert it to an equivalent exponential form.
   (a) $\log_2 8 = 3$ (b) $\log_9 27 = \frac{3}{2}$ (c) $\ln 1 = 0$

3. Simplify the following:

   (a) $\dfrac{\frac{3}{2}\ln 4 - 2\ln 2}{\frac{3}{2}\ln 8}$ (b) $\dfrac{\frac{2}{3}\log 8 + \frac{1}{2}\log 9}{\log 6 + \log 2}$

4. For each of the following equations, sketch its graph.

   (a) $y = \log x$ (b) $y = \log_3 x$

   (c) $y = \ln x$ (d) $y = \log_{\frac{1}{3}} x$

   (e) $y = \log x^2$ (f) $y = 2\log x$

   (g) $y = \log(2x - 1)$

5. For each of the following functions, find its domain and range.

   (a) $f(x) = \mathrm{e}^x + 1$ (b) $f(x) = \mathrm{e}^{x^2}$

   (c) $f(x) = \ln x^2$ (d) $f(x) = 2\ln x$

   (e) $f(x) = \ln(2x - 1)$ (f) $f(x) = \ln(x^2 - 4)$

6. For each of the following equations, find its solution set.

   (a) $\log_{16} x = \frac{1}{2}$ (b) $\log_x 4 = \frac{2}{5}$

   (c) $9^{x-1} = 3^{x+1}$ (d) $x\ln x = \ln x^2$

   (e) $2\log(x - 1) = \log(x^2 - 5)$ (f) $\log(x - 3) = \log(7x - 9) - 1$

7. For each of the following, find x correct to 3 significant figures.

   (a) $10^x = 123.4$ (b) $\mathrm{e}^x = 5678$

   (c) $\log x = -13.57$ (d) $\ln x = 0.0187$

   (e) $1.0375^{4x} = 2$ (f) $\log_x 5 = 2.34$

8. How long will it take money to double if it is invested at 2.275% interest compounded

   (a) annually; (b) quarterly; (c) monthly; (d) continuously

   Give your answer in years correct to two decimal places.

8.3 Differentiation of the Exponential and Logarithmic Functions

Derivative of ln The natural logarithmic function is differentiable on $(0, \infty)$ and its derivative is the reciprocal function, that is,

$$\frac{\mathrm{d}}{\mathrm{d}x}\ln x = \frac{1}{x}, \qquad x > 0.$$

Proof

$$\begin{aligned}
\frac{\mathrm{d}}{\mathrm{d}x}\ln x &= \lim_{h\to 0}\frac{\ln(x+h) - \ln(x)}{h} && \text{Definition of Derivative}\\
&= \lim_{h\to 0}\frac{\ln\left(\frac{x+h}{x}\right)}{h} && \text{Log Property (5)}\\
&= \lim_{h\to 0}\left[\frac{1}{h}\cdot\ln\left(1+\frac{h}{x}\right)\right]\\
&= \lim_{h\to 0}\left[\frac{1}{x}\cdot\frac{x}{h}\cdot\ln\left(1+\frac{h}{x}\right)\right]
\end{aligned}$$

continued on next page

$$\begin{aligned}
\frac{\mathrm{d}}{\mathrm{d}x}\ln x &= \frac{1}{x}\cdot\lim_{h\to 0}\left[\ln\left(1+\frac{h}{x}\right)^{\frac{x}{h}}\right] && \text{Limit Rule (La5s) \& Log Property (6)}\\
&= \frac{1}{x}\ln\left[\lim_{h\to 0}\left(1+\frac{h}{x}\right)^{\frac{x}{h}}\right] && \text{Continuity of ln}\\
&= \frac{1}{x}\ln\left[\lim_{t\to\infty}\left(1+\frac{1}{t}\right)^{t}\right] && \text{Put } t=\frac{x}{h}\\
&= \frac{1}{x}\cdot\ln \mathrm{e} && \text{Definition of e}\\
&= \frac{1}{x} && \text{Log Property (2)}
\end{aligned}$$

□

Example For each of the following y, find $\frac{\mathrm{d}y}{\mathrm{d}x}$.

(1) $y = 2 + 3\ln x$

(2) $y = \sin x \cdot \ln x$

(3) $y = \cos x + \ln x^2$

(4) $y = x\ln 2x$

Solution

(1)
$$\begin{aligned}
\frac{\mathrm{d}y}{\mathrm{d}x} &= \frac{\mathrm{d}}{\mathrm{d}x}(2+3\ln x)\\
&= \frac{\mathrm{d}}{\mathrm{d}x}2 + \frac{\mathrm{d}}{\mathrm{d}x}(3\ln x) && \text{Term by Term Differentiation}\\
&= 0 + 3\cdot\frac{\mathrm{d}}{\mathrm{d}x}\ln x && \text{Derivative of Constant \& Constant Multiple Rule}\\
&= \frac{3}{x} && \text{Derivative of ln}
\end{aligned}$$

(2)
$$\begin{aligned}
\frac{\mathrm{d}y}{\mathrm{d}x} &= \frac{\mathrm{d}}{\mathrm{d}x}(\sin x\cdot\ln x)\\
&= \sin x\cdot\frac{\mathrm{d}}{\mathrm{d}x}\ln x + \ln x\cdot\frac{\mathrm{d}}{\mathrm{d}x}\sin x && \text{Product Rule}\\
&= \frac{\sin x}{x} + \ln x\cdot\cos x && \text{Derivatives of ln \& sin}
\end{aligned}$$

(3)
$$\begin{aligned}
\frac{\mathrm{d}y}{\mathrm{d}x} &= \frac{\mathrm{d}}{\mathrm{d}x}(\cos x + \ln x^2)\\
&= \frac{\mathrm{d}}{\mathrm{d}x}\cos x + \frac{\mathrm{d}}{\mathrm{d}x}\ln x^2 && \text{Term by Term Differentiation}\\
&= -\sin x + \frac{\mathrm{d}}{\mathrm{d}x}(2\ln x) && \text{Derivative of cos \& Log Property (6)}\\
&= -\sin x + 2\cdot\frac{\mathrm{d}}{\mathrm{d}x}\ln x && \text{Constant Multiple Rule}\\
&= -\sin x + \frac{2}{x} && \text{Derivatives of ln}
\end{aligned}$$

$$
\begin{aligned}
(4)\quad \frac{\mathrm{d}y}{\mathrm{d}x} &= \frac{\mathrm{d}}{\mathrm{d}x}(x\ln 2x) \\
&= x\cdot\frac{\mathrm{d}}{\mathrm{d}x}(\ln 2x)+\ln 2x\cdot\frac{\mathrm{d}}{\mathrm{d}x}x && \text{Product Rule} \\
&= x\cdot\frac{\mathrm{d}}{\mathrm{d}x}(\ln 2+\ln x)+\ln 2x && \text{Log Property (3) \& Power Rule} \\
&= x\cdot\left(0+\frac{1}{x}\right)+\ln 2x && \text{Derivatives of Constant \& ln} \\
&= 1+\ln 2x
\end{aligned}
$$

□

Differentiation of Logarithmic functions with other bases To differentiate other logarithmic functions $\log_b x$ where $b \neq \mathrm{e}$, we can use change of base formula:

$$\log_b m = \frac{\ln m}{\ln b}.$$

Example Find the derivative of $y = \log_2(5x^3)$.

Solution

$$
\begin{aligned}
\frac{\mathrm{d}y}{\mathrm{d}x} &= \frac{\mathrm{d}}{\mathrm{d}x}\log_2(5x^3) \\
&= \frac{\mathrm{d}}{\mathrm{d}x}\left(\frac{\ln 5x^3}{\ln 2}\right) && \text{Change of Base Formula} \\
&= \frac{1}{\ln 2}\cdot\frac{\mathrm{d}}{\mathrm{d}x}(\ln 5+3\ln x) && \text{Constant Multiple Rule and Log Properties (3) \& (6)} \\
&= \frac{1}{\ln 2}\left(\frac{\mathrm{d}}{\mathrm{d}x}\ln 5+\frac{\mathrm{d}}{\mathrm{d}x}(3\ln x)\right) && \text{Term by Term Differentiation} \\
&= \frac{1}{\ln 2}\left(0+3\cdot\frac{\mathrm{d}}{\mathrm{d}x}\ln x\right) && \text{Derivative of Constant \& Constant Multiple Rule} \\
&= \frac{1}{\ln 2}\cdot 3\cdot\frac{1}{x} && \text{Derivative of ln} \\
&= \frac{3}{x\ln 2}
\end{aligned}
$$

□

To find a formula for the derivative of the exponential function exp, we use the fact that the functions ln and exp are inverses of each other and we need the following:

Inverse Function Rule $\dfrac{\mathrm{d}x}{\mathrm{d}y} = \dfrac{1}{\dfrac{\mathrm{d}y}{\mathrm{d}x}}$

Explanation More precisely, we have the following result called the *Inverse Function Theorem*.

Let f be a function defined on an open interval (a, b). Suppose that f is differentiable on (a, b) and $f'(x) \neq 0$ for all $x \in (a, b)$. Then on (a, b), the function f is injective, the image of (a, b) under f is an open interval, denoted by (c, d). Moreover, the inverse function f^{-1} is differentiable on (c, d) and

$$(f^{-1})'(\eta) = \frac{1}{f'(\xi)} \quad \text{for all } \eta \in (c, d), \text{ where } \xi = f^{-1}(\eta).$$

If we denote $y = f(x)$, then $x = f^{-1}(y)$ gives the inverse function and we have $\dfrac{\mathrm{d}y}{\mathrm{d}x} = f'(x)$ and $\dfrac{\mathrm{d}x}{\mathrm{d}y} = (f^{-1})'(y)$. The Inverse Function Rule is a compact way to express the relation between the derivatives of f and f^{-1}. Below, we show how to derive the rule:

$$\begin{aligned}
\frac{\mathrm{d}y}{\mathrm{d}x} &= \lim_{\Delta x\to 0}\frac{\Delta y}{\Delta x} && \text{Definition of Derivative}\\
&= \lim_{\Delta x\to 0}\frac{1}{\frac{\Delta x}{\Delta y}} && \text{Note: } \Delta y \neq 0\\
&= \frac{1}{\lim\limits_{\Delta x\to 0}\frac{\Delta x}{\Delta y}} && \text{Continuity of Reciprocal Function}\\
&= \frac{1}{\lim\limits_{\Delta y\to 0}\frac{\Delta x}{\Delta y}} && \text{Continuity of } f\\
&= \frac{1}{\frac{\mathrm{d}x}{\mathrm{d}y}} && \text{Definition of Derivative}
\end{aligned}$$

□

Derivative of exp The exponential function is differentiable on $\mathbb{R}$ and its derivative is the function itself, that is,

$$\frac{\mathrm{d}}{\mathrm{d}x}\mathrm{e}^x = \mathrm{e}^x, \qquad -\infty < x < \infty.$$

Proof Put $y = \mathrm{e}^x$. Then we have $x = \ln y$. From these we get,

$$\begin{aligned}
\frac{\mathrm{d}}{\mathrm{d}x}\mathrm{e}^x &= \frac{\mathrm{d}y}{\mathrm{d}x} && \text{Substitution}\\
&= \frac{1}{\frac{\mathrm{d}x}{\mathrm{d}y}} && \text{Inverse Function Rule}\\
&= \frac{1}{\frac{\mathrm{d}}{\mathrm{d}y}\ln y} && \text{Substitution}\\
&= \frac{1}{\frac{1}{y}} && \text{Derivative of ln}\\
&= y\\
&= \mathrm{e}^x && \text{Substitution}
\end{aligned}$$

□

Example For each of the following y, find $\dfrac{\mathrm{d}y}{\mathrm{d}x}$.

(1) $y = 2\mathrm{e}^x - \dfrac{3}{x}$

(2) $y = x^5\mathrm{e}^x$

(3) $y = \dfrac{\mathrm{e}^x}{\sin x}$

Solution

(1)
$$\begin{aligned}
\frac{dy}{dx} &= \frac{d}{dx}\left(2e^x - \frac{3}{x}\right) \\
&= \frac{d}{dx}(2e^x) - \frac{d}{dx}(3x^{-1}) && \text{Term by Term Differentiation} \\
&= 2 \cdot \frac{d}{dx} e^x - 3 \cdot \frac{d}{dx} x^{-1} && \text{Constant Multiple Rule} \\
&= 2 \cdot e^x - 3 \cdot (-1) \cdot x^{-2} && \text{Derivative of exp \& Power Rule} \\
&= 2e^x + 3x^{-2}
\end{aligned}$$

(2)
$$\begin{aligned}
\frac{dy}{dx} &= \frac{d}{dx}\left(x^5 e^x\right) \\
&= e^x \cdot \frac{d}{dx} x^5 + x^5 \cdot \frac{d}{dx} e^x && \text{Product Rule} \\
&= e^x \cdot 5x^4 + x^5 \cdot e^x && \text{Power Rule \& Derivative of exp} \\
&= (x^5 + 5x^4)\, e^x
\end{aligned}$$

(3)
$$\begin{aligned}
\frac{dy}{dx} &= \frac{d}{dx}\left(\frac{e^x}{\sin x}\right) \\
&= \frac{\sin x \cdot \frac{d}{dx} e^x - e^x \cdot \frac{d}{dx} \sin x}{\sin^2 x} && \text{Quotient Rule} \\
&= \frac{\sin x \cdot e^x - e^x \cdot \cos x}{\sin^2 x} && \text{Derivatives of exp \& sin} \\
&= \frac{e^x (\sin x - \cos x)}{\sin^2 x}
\end{aligned}$$

□

Differentiation of Exponential functions with Bases different from e To differentiate other exponential functions $\exp_b$ where $b \neq e$, we can express b^x in the form $e^{\text{something}}$. Alternatively, we can use a technique called *logarithmic differentiation*. Both methods depends on the *Chain Rule* (see Chapter 9).

To close this chapter, we use the inverse function rule to find the derivative of the arctangent function. The result will be used in the discussion of integration of rational functions. The inverse function rule can also be used to find the derivatives of the functions $\sin^{-1}$ and $\cos^{-1}$ and are left as exercises.

Derivative of $\tan^{-1}$ The arctangent function is differentiable on $\mathbb{R}$ and we have

$$\frac{d}{dx} \tan^{-1} x = \frac{1}{1 + x^2}, \qquad -\infty < x < \infty.$$

Explanation Although the tangent function is not injective, it becomes injective when we restrict the domain to $(-\frac{\pi}{2}, \frac{\pi}{2})$. Note that the range is $\mathbb{R}$. The inverse of the function $\tan : (-\frac{\pi}{2}, \frac{\pi}{2}) \longrightarrow \mathbb{R}$ is called the *arctangent function*, denoted by arctan or $\tan^{-1}$. Thus $\tan^{-1}$ is the function from $\mathbb{R}$ into $(-\frac{\pi}{2}, \frac{\pi}{2})$ such that

$$\tan(\tan^{-1} x) = x \quad \text{for all } x \in \mathbb{R} \qquad \text{and} \qquad \tan^{-1}(\tan x) = x \quad \text{for all } x \in (-\tfrac{\pi}{2}, \tfrac{\pi}{2}).$$

Proof Put $y = \tan^{-1} x$. Then we have $x = \tan y$. From these we get

$$\begin{aligned}
\frac{\mathrm{d}}{\mathrm{d}x}\tan^{-1}x &= \frac{\mathrm{d}y}{\mathrm{d}x} && \text{Substitution} \\
&= \frac{1}{\frac{\mathrm{d}x}{\mathrm{d}y}} && \text{Inverse Function Rule} \\
&= \frac{1}{\frac{\mathrm{d}}{\mathrm{d}y}\tan y} && \text{Substitution} \\
&= \frac{1}{\sec^2 y} && \text{Derivative of tan} \\
&= \frac{1}{1+\tan^2 y} && \text{Identity (Py1)} \\
&= \frac{1}{1+x^2} && \text{Substitution}
\end{aligned}$$

□

Exercise 8.3

1. For each of the following y, find $\frac{\mathrm{d}y}{\mathrm{d}x}$.
   (a) $y = 2x^3 - 4\mathrm{e}^x - 5$ (b) $y = \ln x - 1$
   (c) $y = \mathrm{e}^x + \ln x$ (d) $y = x^2 + \ln x^2$
   (e) $y = \mathrm{e}^x + \sqrt{x}$ (f) $y = \ln\sqrt{x} - 1$
   (g) $y = \mathrm{e}^x \sin x$ (h) $y = \cos x \ln x$
   (i) $y = (x^2+1)\mathrm{e}^x$ (j) $y = (x^2+1)\ln x$
   (k) $y = \frac{\mathrm{e}^x}{\cos x}$ (l) $y = \frac{\ln x}{\sin x}$
   (m) $y = \frac{x^3+3x^2+6x-2}{\mathrm{e}^x}$ (n) $y = \frac{x^3+3x^2+6x-2}{\ln x}$
2. For each of the following f, find $f'(a)$ for the given a.
   (a) $f(x) = \mathrm{e}^x \tan x, \quad a = 0$
   (b) $f(x) = \frac{\ln x}{x^2+1}, \quad a = 1$
3. For each of the following f, find $\frac{\mathrm{d}^2y}{\mathrm{d}x^2}$.
   (a) $y = x^2 + x - 1 - \mathrm{e}^x$
   (b) $y = 1 - \ln x - \frac{1}{x}$
4. Use the inverse function rule to prove the following:
   (a) $\frac{\mathrm{d}}{\mathrm{d}x}\sin^{-1}x = \frac{1}{\sqrt{1-x^2}}, \quad -1 < x < 1$
   (b) $\frac{\mathrm{d}}{\mathrm{d}x}\cos^{-1}x = -\frac{1}{\sqrt{1-x^2}}, \quad -1 < x < 1$

   Note: The functions $\sin^{-1}$ and $\cos^{-1}$ are the inverses of the injective functions $\sin : [-\frac{\pi}{2}, \frac{\pi}{2}] \longrightarrow [-1, 1]$ and $\cos : [0, \pi] \longrightarrow [-1, 1]$ respectively.

Chapter 9

More Differentiation

9.1 Chain rule

Up to this stage, we know how to differentiate "simple" functions like the following:

- $f(x) = x^5 + 1$
- $f(x) = \frac{x-1}{x+1}$
- $f(x) = \sin x$
- $f(x) = \mathrm{e}^x + 2\tan x$
- $f(x) = \frac{\ln x}{\cos x} - \frac{\mathrm{e}^x}{x^2+1}$

using simple rules for differentiation and formulas derived in the last few chapters. But for more "complicated" functions, like the following:

- $f(x) = \sin(x^2)$
- $f(x) = \mathrm{e}^{x^2+1}$
- $f(x) = \ln(1+2x)$

we need the *chain rule*. It is one of the most important rules for finding derivatives, used for differentiating composite functions.

Chain Rule $\quad \dfrac{\mathrm{d}y}{\mathrm{d}x} = \dfrac{\mathrm{d}y}{\mathrm{d}u} \cdot \dfrac{\mathrm{d}u}{\mathrm{d}x}$

Explanation More precisely, we have the following result for differentiation of composition of functions.

Let f be a function that is differentiable on an open interval (a, b). Let g be a function that is differentiable on an open interval containing the image of (a, b) under f. Then the composition $g \circ f$ is differentiable on (a, b). Moreover, we have

$$(g \circ f)'(\xi) = g'(f(\xi)) \cdot f'(\xi) \quad \text{for all } \xi \in (a, b).$$

If we denote $u = f(x)$ and denote $y = g(u)$, then $y = g(f(x))$ is a function of x. Note that $(g \circ f)' = \frac{\mathrm{d}y}{\mathrm{d}x}$, $g' = \frac{\mathrm{d}y}{\mathrm{d}u}$ and $f' = \frac{\mathrm{d}u}{\mathrm{d}x}$. The Chain Rule is a compact way to express the relation between the derivatives of $g \circ f$, g and f. Below we show how to derive the rule (with an additional assumption):

$$
\begin{aligned}
\frac{\mathrm{d}y}{\mathrm{d}x} &= \lim_{\Delta x\to 0}\frac{\Delta y}{\Delta x} \\
&= \lim_{\Delta x\to 0}\left(\frac{\Delta y}{\Delta u}\cdot\frac{\Delta u}{\Delta x}\right) && \text{Assume } \Delta u \neq 0 \\
&= \lim_{\Delta x\to 0}\frac{\Delta y}{\Delta u}\cdot\lim_{\Delta x\to 0}\frac{\Delta u}{\Delta x} && \text{Limit Rule (L.5)} \\
&= \lim_{\Delta u\to 0}\frac{\Delta y}{\Delta u}\cdot\lim_{\Delta x\to 0}\frac{\Delta u}{\Delta x} && \text{Continuity of } f \\
&= \frac{\mathrm{d}y}{\mathrm{d}u}\cdot\frac{\mathrm{d}u}{\mathrm{d}x}
\end{aligned}
$$

Example Find $\dfrac{\mathrm{d}}{\mathrm{d}x}(x^2+5)^3$

(1) without using chain rule;

(2) using chain rule.

Solution

(1)
$$
\begin{aligned}
\frac{\mathrm{d}}{\mathrm{d}x}(x^2+5)^3 &= \frac{\mathrm{d}}{\mathrm{d}x}(x^6+15x^4+75x^2+125) && \text{Rewrite the function} \\
&= 6x^5+15\cdot 4x^3+75\cdot 2x+0 && \text{Term by Term Differentiation,} \\
& && \text{Constant Multiple Rule \& Power Rule} \\
&= 6x^5+60x^3+150x
\end{aligned}
$$

(2) Put $u = x^2 + 5$ and put $y = u^3$. Then we have $y = (x^2 + 5)^3$. From these we get

$$
\begin{aligned}
\frac{\mathrm{d}}{\mathrm{d}x}(x^2+5)^3 &= \frac{\mathrm{d}y}{\mathrm{d}x} && \text{Substitution} \\
&= \frac{\mathrm{d}y}{\mathrm{d}u}\cdot\frac{\mathrm{d}u}{\mathrm{d}x} && \text{Chain Rule} \\
&= \frac{\mathrm{d}}{\mathrm{d}u}u^3\cdot\frac{\mathrm{d}}{\mathrm{d}x}(x^2+5) && \text{Substitution} \\
&= 3u^2\cdot(2x+0) && \text{Power Rule and Term by Term Differentiation} \\
&= 3(x^2+5)^2\cdot 2x && \text{Substitution} \\
&= 6x\,(x^2+5)^2
\end{aligned}
$$

□

Remark

- It is straightforward to check that the above two results are the same.
- If we change the function to $(x^2+5)^{\frac{1}{3}}$, we can't apply the first method but can still apply the second method which makes use of the chain rule together with the power rule.

We can combine the chain rule with any formula to get a more general formula. In the table below, the general form gives the derivative of $g \circ f$ where g is a power function, the sine function etc. and f is a differentiable function such that $g \circ f$ is defined (for example, in order that $\ln[f(x)]$ be defined, we have to assume that f is

positive). These formulas will be referred as the *Chain Rule & Power Rule*, *Chain Rule & Derivative of* sin etc. For the Power Rule, we have seen that it is true if r is an integer or a rational number in the form $n + \frac{1}{2}$ where n is an integer. In fact, it is true for all real numbers r. We will prove this result (called the *General Power Rule*) later using *logarithmic differentiation* which is based on the Chain Rule & Derivative of ln.

Simple Form	General Form
$\frac{\mathrm{d}}{\mathrm{d}x}x^r = rx^{r-1}$	$\frac{\mathrm{d}}{\mathrm{d}x}[f(x)]^r = r[f(x)]^{r-1} \cdot \frac{\mathrm{d}}{\mathrm{d}x}f(x)$
$\frac{\mathrm{d}}{\mathrm{d}x}\sin x = \cos x$	$\frac{\mathrm{d}}{\mathrm{d}x}\sin[f(x)] = \cos[f(x)] \cdot \frac{\mathrm{d}}{\mathrm{d}x}f(x)$
$\frac{\mathrm{d}}{\mathrm{d}x}\cos x = -\sin x$	$\frac{\mathrm{d}}{\mathrm{d}x}\cos[f(x)] = -\sin[f(x)] \cdot \frac{\mathrm{d}}{\mathrm{d}x}f(x)$
$\frac{\mathrm{d}}{\mathrm{d}x}\tan x = \sec^2 x$	$\frac{\mathrm{d}}{\mathrm{d}x}\tan[f(x)] = \sec^2[f(x)] \cdot \frac{\mathrm{d}}{\mathrm{d}x}f(x)$
$\frac{\mathrm{d}}{\mathrm{d}x}\mathrm{e}^x = \mathrm{e}^x$	$\frac{\mathrm{d}}{\mathrm{d}x}\mathrm{e}^{f(x)} = \mathrm{e}^{f(x)} \cdot \frac{\mathrm{d}}{\mathrm{d}x}f(x)$
$\frac{\mathrm{d}}{\mathrm{d}x}\ln x = \frac{1}{x}$	$\frac{\mathrm{d}}{\mathrm{d}x}\ln[f(x)] = \frac{1}{f(x)} \cdot \frac{\mathrm{d}}{\mathrm{d}x}f(x)$

Proof We give the proof for the 1st, 2nd and 6th formulas. The proofs of the rest are left as exercises.

(1) Put $u = f(x)$ and put $y = u^r$. Then we have $y = [f(x)]^r$. From these we get

$$\begin{aligned}
\frac{\mathrm{d}}{\mathrm{d}x}[f(x)]^r &= \frac{\mathrm{d}y}{\mathrm{d}x} && \text{Substitution} \\
&= \frac{\mathrm{d}y}{\mathrm{d}u} \cdot \frac{\mathrm{d}u}{\mathrm{d}x} && \text{Chain Rule} \\
&= \frac{\mathrm{d}}{\mathrm{d}u}u^r \cdot \frac{\mathrm{d}u}{\mathrm{d}x} && \text{Substitution} \\
&= ru^{r-1} \cdot \frac{\mathrm{d}u}{\mathrm{d}x} && \text{Power Rule} \\
&= r[f(x)]^{r-1} \cdot \frac{\mathrm{d}}{\mathrm{d}x}f(x) && \text{Substitution}
\end{aligned}$$

(2) Put $u = f(x)$ and put $y = \sin u$. Then we have $y = \sin[f(x)]$. From these we get

$$\begin{aligned}
\frac{\mathrm{d}}{\mathrm{d}x}\sin[f(x)] &= \frac{\mathrm{d}y}{\mathrm{d}x} && \text{Substitution} \\
&= \frac{\mathrm{d}y}{\mathrm{d}u} \cdot \frac{\mathrm{d}u}{\mathrm{d}x} && \text{Chain Rule} \\
&= \frac{\mathrm{d}}{\mathrm{d}u}\sin u \cdot \frac{\mathrm{d}u}{\mathrm{d}x} && \text{Substitution} \\
&= \cos u \cdot \frac{\mathrm{d}u}{\mathrm{d}x} && \text{Derivative of sin} \\
&= \cos[f(x)] \cdot \frac{\mathrm{d}}{\mathrm{d}x}f(x) && \text{Substitution}
\end{aligned}$$

(6) Put $u = f(x)$ and put $y = \ln u$. Then we have $y = \ln[f(x)]$. From these we get

$$\begin{aligned}
\frac{\mathrm{d}}{\mathrm{d}x}\ln[f(x)] &= \frac{\mathrm{d}y}{\mathrm{d}x} && \text{Substitution} \\
&= \frac{\mathrm{d}y}{\mathrm{d}u}\cdot\frac{\mathrm{d}u}{\mathrm{d}x} && \text{Chain Rule} \\
&= \frac{\mathrm{d}}{\mathrm{d}u}\ln u\cdot\frac{\mathrm{d}u}{\mathrm{d}x} && \text{Substitution} \\
&= \frac{1}{u}\cdot\frac{\mathrm{d}u}{\mathrm{d}x} && \text{Derivative of ln} \\
&= \frac{1}{f(x)}\cdot\frac{\mathrm{d}}{\mathrm{d}x}f(x) && \text{Substitution}
\end{aligned}$$

□

Example For each of the following y, find $\frac{\mathrm{d}y}{\mathrm{d}x}$.

(1) $y = \sin(x^2 + 1)$

(2) $y = \mathrm{e}^{x^2+2}$

(3) $y = \ln(x^2 - 3)$

(4) $y = \mathrm{e}^{x^2+\tan x^2}$

(5) $y = \ln[\sin^2(2x + 3)]$

(6) $y = \dfrac{1}{(x^2 + 3)^4}$

(7) $y = \mathrm{e}^{x+1}\ln(x^2 + 1)$

Solution

(1)
$$\begin{aligned}
\frac{\mathrm{d}y}{\mathrm{d}x} &= \frac{\mathrm{d}}{\mathrm{d}x}\sin(x^2+1) \\
&= \cos(x^2+1)\cdot\frac{\mathrm{d}}{\mathrm{d}x}(x^2+1) && \text{Chain Rule \& Derivative of sin} \\
&= 2x\cos(x^2+1) && \text{Term by Term Differentiation and Power Rule}
\end{aligned}$$

(2)
$$\begin{aligned}
\frac{\mathrm{d}y}{\mathrm{d}x} &= \frac{\mathrm{d}}{\mathrm{d}x}\mathrm{e}^{x^2+2} \\
&= \mathrm{e}^{x^2+2}\cdot\frac{\mathrm{d}}{\mathrm{d}x}(x^2+2) && \text{Chain Rule \& Derivative of exp} \\
&= 2x\,\mathrm{e}^{x^2+2} && \text{Term by Term Differentiation and Power Rule}
\end{aligned}$$

(3)
$$\begin{aligned}
\frac{\mathrm{d}y}{\mathrm{d}x} &= \frac{\mathrm{d}}{\mathrm{d}x}\ln(x^2-3) \\
&= \frac{1}{x^2-3}\cdot\frac{\mathrm{d}}{\mathrm{d}x}(x^2-3) && \text{Chain Rule \& Derivative of ln} \\
&= \frac{2x}{x^2-3} && \text{Term by Term Differentiation and Power Rule}
\end{aligned}$$

(4)
$$\begin{aligned}
\frac{dy}{dx} &= \frac{d}{dx}e^{x^2+\tan x^2} \\
&= e^{x^2+\tan x^2}\cdot\frac{d}{dx}(x^2+\tan x^2) && \text{Chain Rule \& Derivative of exp} \\
&= e^{x^2+\tan x^2}\cdot\left(\frac{d}{dx}x^2+\frac{d}{dx}\tan x^2\right) && \text{Term by Term Differentiation} \\
&= e^{x^2+\tan x^2}\cdot\left(2x+\sec^2 x^2\cdot\frac{d}{dx}x^2\right) && \text{Power Rule and Chain Rule \& Derivative of tan} \\
&= e^{x^2+\tan x^2}\cdot(2x+\sec^2 x^2\cdot 2x) && \text{Power Rule} \\
&= 2x\,(1+\sec^2 x^2)\,e^{x^2+\tan x^2}
\end{aligned}$$

(5)
$$\begin{aligned}
\frac{dy}{dx} &= \frac{d}{dx}\ln[\sin^2(2x+3)] \\
&= \frac{1}{\sin^2(2x+3)}\cdot\frac{d}{dx}\sin^2(2x+3) && \text{Chain Rule \& Derivative of ln} \\
&= \frac{1}{\sin^2(2x+3)}\cdot\frac{d}{dx}[\sin(2x+3)]^2 && \text{Rewrite } \sin^2 u \text{ as } (\sin u)^2 \\
&= \frac{1}{\sin^2(2x+3)}\cdot 2\sin(2x+3)\cdot\frac{d}{dx}\sin(2x+3) && \text{Chain Rule \& Power Rule} \\
&= \frac{1}{\sin^2(2x+3)}\cdot 2\sin(2x+3)\cdot\cos(2x+3)\cdot\frac{d}{dx}(2x+3) && \text{Chain Rule \& Derivative of sin} \\
&= \frac{4\cos(2x+3)}{\sin(2x+3)} && \text{Term by Term Differentiation and Power Rule}
\end{aligned}$$

Remark In the above solution, we apply the chain rule thrice. Alternatively, we may use a property of logarithm and apply the chain rule twice.

Alternative solution

$$\begin{aligned}
\frac{dy}{dx} &= \frac{d}{dx}\ln[\sin^2(2x+3)] \\
&= \frac{d}{dx}\ln[\sin(2x+3)]^2 && \text{Rewrite the function} \\
&= \frac{d}{dx}2\ln[\sin(2x+3)] && \text{Log Property (6)} \\
&= 2\cdot\frac{d}{dx}\ln[\sin(2x+3)] && \text{Constant Multiple Rule} \\
&= 2\cdot\frac{1}{\sin(2x+3)}\cdot\frac{d}{dx}\sin(2x+3) && \text{Chain Rule \& Derivative of ln} \\
&= \frac{2}{\sin(2x+3)}\cdot\cos(2x+3)\cdot\frac{d}{dx}(2x+3) && \text{Chain Rule \& Derivative of sin} \\
&= \frac{4\cos(2x+3)}{\sin(2x+3)} && \text{Term by Term Differentiation and Power Rule}
\end{aligned}$$

(6)
$$\begin{aligned}
\frac{dy}{dx} &= \frac{d}{dx}\frac{1}{(x^2+3)^4} \\
&= \frac{d}{dx}(x^2+3)^{-4} && \text{Rewrite the function} \\
&= (-4)\cdot(x^2+3)^{-5}\cdot\frac{d}{dx}(x^2+3) && \text{Chain Rule \& Power Rule} \\
&= \frac{-4}{(x^2+3)^5}\cdot 2x && \text{Term by Term Differentiation and Power Rule} \\
&= \frac{-8x}{(x^2+3)^5}
\end{aligned}$$

Remark If we apply the quotient rule in the first step, we still have to apply the chain rule in a later step.

(7)
$$\begin{aligned}
\frac{dy}{dx} &= \frac{d}{dx}\left(e^{x+1}\cdot\ln(x^2+1)\right) \\
&= e^{x+1}\cdot\frac{d}{dx}\ln(x^2+1)+\ln(x^2+1)\cdot\frac{d}{dx}e^{x+1} && \text{Product Rule} \\
&= e^{x+1}\cdot\frac{1}{x^2+1}\cdot\frac{d}{dx}(x^2+1)+\ln(x^2+1)\cdot e^{x+1}\cdot\frac{d}{dx}(x+1) && \text{Chain Rule \& Derivative of ln and Chain Rule \& Derivative of exp} \\
&= \frac{2x\,e^{x+1}}{x^2+1}+e^{x+1}\ln(x^2+1) && \text{Term by Term Differentiation and Power Rule}
\end{aligned}$$

□

Remark Instead of using the General Form given in the table on page 209, some authors use the chain rule directly by writing down the expression for u. Below we redo (1) and (2) in the last example using such notations.

(1)
$$\begin{aligned}
\frac{dy}{dx} &= \frac{d}{dx}\sin(x^2+1) \\
&= \frac{d}{d(x^2+1)}\sin(x^2+1)\cdot\frac{d}{dx}(x^2+1) && \text{Chain Rule (replace } u \text{ by } x^2+1) \\
&= 2x\cos(x^2+1) && \text{Term by Term Differentiation, Power Rule \& Derivative of sin}
\end{aligned}$$

(2)
$$\begin{aligned}
\frac{dy}{dx} &= \frac{d}{dx}e^{x^2+2} \\
&= \frac{d}{d(x^2+2)}e^{x^2+2}\cdot\frac{d}{dx}(x^2+2) && \text{Chain Rule (replace } u \text{ by } x^2+2) \\
&= 2x\,e^{x^2+2} && \text{Term by Term Differentiation, Power Rule \& Derivative of exp}
\end{aligned}$$

Logarithmic Differentiation Suppose that y is a differentiable function of x and that y is always positive. Then the composition $\ln y$ is a differentiable function of x. Moreover, by the Chain Rule & Derivative of ln, we have

$$\frac{d}{dx}\ln y = \frac{1}{y}\cdot\frac{dy}{dx}. \tag{9.1.1}$$

Using properties of logarithms, we may be able to find $\frac{d}{dx}\ln y$. In this case, we can then find $\frac{dy}{dx}$ using (9.1.1. This method is called *logarithmic differentiation*.

Below we will apply logarithmic differentiation to proof the General Power Rule and do an example to illustrate how to find the *derivatives of exponential functions with bases different from* e.

General Power Rule Let r be a real number. Then the power function x^r is differentiable on $(0, \infty)$. Moreover, we have

$$\frac{d}{dx}x^r = rx^{r-1}, \qquad x > 0.$$

Explanation We can use logarithmic differentiation because $\ln x^r = r\ln x$ can be differentiated easily.

Proof Put $y = x^r$. Taking natural logarithm and using Log Property (6), we get

$$\ln y = r\ln x.$$

Differentiating both sides with respect to x, we get

$$\begin{aligned}
\frac{d}{dx}\ln y &= \frac{d}{dx}(r\ln x) \\
\frac{1}{y}\cdot\frac{dy}{dx} &= r\cdot\frac{1}{x} && \text{Chain Rule \& Derivative of ln, Constant Multiple Rule and Derivative of ln} \\
\frac{dy}{dx} &= y\cdot\frac{r}{x} \\
&= x^r\cdot\frac{r}{x} && \text{Substitution} \\
&= rx^{r-1}
\end{aligned}$$

□

Example Find $\frac{d}{dx}5^{x^2+\cos x}$.

Explanation The given function is in the form $b^{f(x)}$ where $b \neq e$ and f is a differentiable function such that f' can be found easily. Its derivative can be found by the following two methods:

(*Method 1*) Express $b^{f(x)}$ in the form $e^{g(x)}$ and then apply Chain Rule & Derivative of exp.

(*Method 2*) Use logarithmic differentiation: note that $\ln b^{f(x)} = f(x)\ln b$ can be differentiated easily.

Solution 1 Note that $5^u = e^{\ln 5^u} = e^{u\ln 5}$ by Log Properties (8) and (6). Therefore, we have

$$\begin{aligned}
\frac{d}{dx}5^{x^2+\cos x} &= \frac{d}{dx}e^{(x^2+\cos x)\ln 5} && \text{Rewrite the function} \\
&= e^{(x^2+\cos x)\ln 5}\cdot\frac{d}{dx}(x^2+\cos x)\cdot\ln 5 && \text{Chain Rule \& Derivative of exp} \\
&= e^{(x^2+\cos x)\ln 5}\cdot\ln 5\cdot(2x-\sin x) && \text{Constant Multiple Rule, Term by Term Differentiation, Power Rule and Derivative of cos} \\
&= (2x-\sin x)\,5^{x^2+\cos x}\ln 5 && \text{Rewrite the function}
\end{aligned}$$

Solution 2

(Step 1) Put $y = 5^{x^2+\cos x}$.

(Step 2) Taking natural logarithm, we get

$$\ln y = (x^2 + \cos x)\ln 5.$$

(Step 3) Differentiating both sides with respect to x, we get

$$\begin{aligned}
\frac{d}{dx}\ln y &= \frac{d}{dx}(x^2 + \cos x)\ln 5 \\
\frac{1}{y}\cdot\frac{dy}{dx} &= \ln 5\cdot(2x - \sin x) && \text{Chain Rule \& Derivative of ln, Constant Multiple Rule, Term by Term Differentiation, Power Rule and Derivative of cos} \\
\frac{dy}{dx} &= y\,(2x - \sin x)\ln 5 \\
&= (2x - \sin x)\,5^{x^2+\cos x}\ln 5 && \text{Substitution}
\end{aligned}$$

□

Exercise 9.1

1. Find $\frac{dy}{dx}$ for the following:

(a)	$y = (2x^3 + 5x^2)^6$	(b)	$y = \sqrt{9 + 4x}$
(c)	$y = \frac{\sqrt{4x^2+5}}{3}$	(d)	$y = \cos 5x$
(e)	$y = \sin(6x - 7)$	(f)	$y = \sin^4 x$
(g)	$y = \cos^5(6x - 7)$	(h)	$y = 4x \sin 3x$
(i)	$y = \tan(8x^3 + 1)$	(j)	$y = \frac{\tan x}{x+2}$
(k)	$y = 2e^{3x} + 4x - 5$	(l)	$y = x\,e^{x^2}$
(m)	$y = \frac{x^2}{e^x}$	(n)	$y = \ln 8x$
(o)	$y = \ln(5 - 2x)$	(p)	$y = \ln(1 - x^2)$
(q)	$y = \ln\sqrt{2x + 11}$	(r)	$y = 3x \ln x$
(s)	$y = \ln(\ln x)$	(t)	$y = e^{x^2}\ln x$
(u)	$y = e^{\tan x}$	(v)	$y = \tan(e^x)$
(w)	$y = \sin(e^{5x})$	(x)	$y = e^{\sin 5x}$
(y)	$y = \cos[\ln(4x^2 + 9)]$	(z)	$y = \ln[\cos(4x^2 + 9)]$

2. Use logarithmic differentiation to find $\frac{dy}{dx}$ for the following:

(a)	$y = 2^{x^2+1}$	(b)	$y = x^x$
(c)	$y = (\sin x)^{\cos x}$	(d)	$y = \frac{(2x+1)(3x+4)^5}{(x^2+7)^8}$

3. Ecologists estimate that when the population of a certain city is x thousand persons, the average level L of carbon monoxide in the air above the city will be L ppm (parts per million), where $L = 10+0.4x+0.001x^2$. The population of the city is estimated to be $x = 345 + 22t + 0.5t^2$ thousand persons t years from the present.
   (a) Find the rate of change of carbon monoxide with respect to the population of the city.
   (b) Find the time rate of change of the population when $t = 3$.
   (c) How fast (with respect to time) is the carbon monoxide level changing at time $t = 3$?

4. Suppose that f and g are differentiable functions such that $f(1) = 2$, $f'(1) = 3$, $f'(5) = 4$, $g(1) = 5$, $g'(1) = 6$, $g'(2) = 7$ and $g(5) = 8$. Find $\frac{\mathrm{d}}{\mathrm{d}x} f(g(x))\Big|_{x=1}$.

9.2 Implicit Differentiation

Implicit differentiation is a technique for differentiating functions that are not given in the usual form $y = f(x)$. We use the following example (Solution 2) to illustrate the general procedure.

Example Find the slope of the line tangent to the circle

$$x^2 + y^2 = 4 \tag{9.2.1}$$

at the point $(\sqrt{2}, \sqrt{2}\,)$.

Explanation There are two methods to find the slope.

(1) The first method is to rewrite the equation of the circle (in fact, semi-circle) in the form $y = f(x)$. The slope at the given point is $f'(\sqrt{2}\,)$. From (9.2.1), solving y in terms of x, we get $y = \pm\sqrt{4 - x^2}$. But this doesn't give a function since certain value of x (say $x = 1$) gives two values of y. Note that the point $(\sqrt{2}, \sqrt{2}\,)$ lies on the upper semi-circle. To consider the required slope, we take $y = \sqrt{4 - x^2}$.

(2) The second method is to differentiate both sides of Equation (9.2.1) with respect to x. For the left-side, to find $\frac{\mathrm{d}}{\mathrm{d}x} y^2$ we can use the Chain Rule & Power Rule by treating $y = f(x)$ as a function of x.

Solution 1 From (9.2.1) and noting that the point $(\sqrt{2}, \sqrt{2}\,)$ lies on the upper circle, we get $y = \sqrt{4 - x^2}$. Differentiating, we get

$$\begin{aligned}
\frac{\mathrm{d}y}{\mathrm{d}x} &= \frac{\mathrm{d}}{\mathrm{d}x}(4 - x^2)^{\frac{1}{2}} \\
&= \frac{1}{2} \cdot (4 - x^2)^{-\frac{1}{2}} \cdot \frac{\mathrm{d}}{\mathrm{d}x}(4 - x^2) && \text{Chain Rule \& Power Rule} \\
&= \frac{1}{2 \cdot (4 - x^2)^{\frac{1}{2}}} \cdot (-2x) && \text{Term by Term Differentiation, and Power Rule} \\
&= \frac{-x}{\sqrt{4 - x^2}}.
\end{aligned}$$

The slope of the tangent at $(\sqrt{2}, \sqrt{2}\,)$ is

$$\frac{\mathrm{d}y}{\mathrm{d}x}\Big|_{x=\sqrt{2}} = \frac{-\sqrt{2}}{\sqrt{2}} = -1.$$

□

Solution 2 Differentiate both sides of (9.2.1) with respect to x, we get

$$\begin{aligned}
\frac{\mathrm{d}}{\mathrm{d}x}(x^2 + y^2) &= \frac{\mathrm{d}}{\mathrm{d}x}(4) \\
\frac{\mathrm{d}}{\mathrm{d}x}x^2 + \frac{\mathrm{d}}{\mathrm{d}x}y^2 &= 0 && \text{Term by Term Differentiation and Derivative of Constant} \\
2x + 2y\frac{\mathrm{d}y}{\mathrm{d}x} &= 0 && \text{Power Rule and Chain Rule \& Power Rule}
\end{aligned}$$

Solving for $\frac{dy}{dx}$, we get

$$\frac{dy}{dx} = -\frac{x}{y}.$$

At $(\sqrt{2}, \sqrt{2})$, the slope of the tangent is

$$\left.\frac{dy}{dx}\right|_{(\sqrt{2},\sqrt{2})} = -\frac{\sqrt{2}}{\sqrt{2}} = -1.$$

□

In general, although an equation of the form

$$F(x, y) = 0 \tag{9.2.2}$$

usually defines y as a function of x implicitly, it may be difficult to express y in terms of x explicitly. By treating y as a function of x, the left-side of (9.2.2) becomes a function of x. To find $\frac{dy}{dx}$, we may differentiate both sides of (9.2.2) with respect to x and then solve for $\frac{dy}{dx}$ (in terms of x and y). This method is called *implicit differentiation*.

Example Use implicit differentiation to find $\frac{dy}{dx}$ given that $x^3 + 4xy^2 - 7 = y^3$.

Explanation To apply implicit differentiation, we assume that y is a function of x. Thus, $4xy^2$ is a function of x and y^3 is a function of x. Moreover, $4xy^2$ can be treated as a product of functions of x.

Solution Differentiate both sides of the given equation with respect to x, we get

$$\begin{aligned}
\frac{d}{dx}(x^3 + 4xy^2 - 7) &= \frac{d}{dx}(y^3) \\
\frac{d}{dx}x^3 + \frac{d}{dx}(4xy^2) - \frac{d}{dx}7 &= 3y^2 \cdot \frac{dy}{dx} && \text{Term by Term Differentiation and Chain Rule \& Power Rule} \\
3x^2 + 4x \cdot \frac{d}{dx}(y^2) + y^2 \cdot \frac{d}{dx}(4x) - 0 &= 3y^2\frac{dy}{dx} && \text{Power Rule and Product Rule} \\
3x^2 + 4x \cdot \left(2y \cdot \frac{dy}{dx}\right) + y^2 \cdot 4 &= 3y^2\frac{dy}{dx} && \text{Chain Rule \& Power Rule, Constant Multiple Rule and Power Rule} \\
3x^2 + 8xy\frac{dy}{dx} + 4y^2 &= 3y^2\frac{dy}{dx}.
\end{aligned}$$

Solving for $\frac{dy}{dx}$, we get

$$\begin{aligned}
3x^2 + 4y^2 &= 3y^2\frac{dy}{dx} - 8xy\frac{dy}{dx} && \text{Collect similar terms} \\
3x^2 + 4y^2 &= \frac{dy}{dx} \cdot (3y^2 - 8xy) && \text{Extract common factor} \\
\frac{dy}{dx} &= \frac{3x^2 + 4y^2}{3y^2 - 8xy}.
\end{aligned}$$

□

Caution $\frac{dy}{dx}(3y^2 - 8xy)$ and $\frac{d}{dx}(3y^2 - 8xy)$ are different. The first expression is the product of $\frac{dy}{dx}$ and $3y^2 - 8xy$ whereas the second one is the derivative of $3y^2 - 8xy$.

Example Find $\dfrac{dy}{dx}$ given that $y \ln x = x e^y - 1$.

Explanation The question is to use implicit differentiation to find $\frac{dy}{dx}$ (it is difficult or even impossible to solve y as a function of x explicitly).

Solution Differentiating both sides of the given equation with respect to x, we get

$$\begin{aligned}
\frac{d}{dx}\big(y \ln x\big) &= \frac{d}{dx}\big(x e^y - 1\big) \\
y \cdot \frac{d}{dx} \ln x + \ln x \cdot \frac{dy}{dx} &= \left(x \cdot \frac{d}{dx} e^y + e^y \cdot \frac{d}{dx} x\right) - \frac{d}{dx} 1 && \text{Term by Term Differentiation} \\
y \cdot \frac{1}{x} + \ln x \cdot \frac{dy}{dx} &= x \cdot e^y \cdot \frac{dy}{dx} + e^y && \text{Derivatives of ln \& exp and Power Rule.}
\end{aligned}$$

Solving for $\frac{dy}{dx}$, we get

$$\begin{aligned}
x \ln x \frac{dy}{dx} - x^2 e^y \frac{dy}{dx} &= x e^y - y && \text{Multiply by } x \text{ and collect similar terms} \\
\frac{dy}{dx} \cdot (x \ln x - x^2 e^y) &= x e^y - y && \text{Extract common factor} \\
\frac{dy}{dx} &= \frac{x e^y - y}{x \ln x - x^2 e^y}.
\end{aligned}$$

□

Example Find the slope of the curve with equation

$$x \sin y + \cos y^2 = 1$$

at the point $(1, 0)$.

Explanation The required slope is $\left.\frac{dy}{dx}\right|_{(1,0)}$. We use implicit differentiation to find $\frac{dy}{dx}$ and then substitute $(x, y) = (1, 0)$.

Solution Differentiating both sides of the given equation with respect to x, we get

$$\begin{aligned}
\frac{d}{dx}\big(x \sin y + \cos y^2\big) &= \frac{d}{dx} 1 \\
\frac{d}{dx}(x \sin y) + \frac{d}{dx} \cos y^2 &= 0 && \text{Term by Term Differentiation and Derivative of Constant} \\
\left(x \cdot \frac{d}{dx} \sin y + \sin y \cdot \frac{d}{dx} x\right) + (-\sin y^2) \cdot \frac{d}{dx} y^2 &= 0 && \text{Product Rule and Chain Rule \& Derivative of cos} \\
(\dagger) \qquad \left(x \cdot \cos y \cdot \frac{dy}{dx} + \sin y \cdot 1\right) - \sin y^2 \cdot 2y \cdot \frac{dy}{dx} &= 0 && \text{Derivative of sin, Power Rule and Chain Rule \& Power Rule.}
\end{aligned}$$

Solving for $\frac{dy}{dx}$, we get

$$\begin{aligned}
\sin y &= 2y \sin y^2 \frac{dy}{dx} - x \cos y \frac{dy}{dx} && \text{Collect similar terms} \\
\sin y &= \frac{dy}{dx} \cdot (2y \sin y^2 - x \cos y) && \text{Extract common factor} \\
\frac{dy}{dx} &= \frac{\sin y}{2y \sin y^2 - x \cos y}.
\end{aligned}$$

The slope of the curve at $(1,0)$ is

$$\left.\frac{dy}{dx}\right|_{(1,0)} = \frac{\sin 0}{0 - \cos 0} = 0.$$

□

Remark Alternatively, we can substitute $(x,y) = (1,0)$ into (†) to get

$$\cos 0 \cdot \left.\frac{dy}{dx}\right|_{(1,0)} + \sin 0 \cdot \left.\frac{dy}{dx}\right|_{(1,0)} = 0 \cdot \left.\frac{dy}{dx}\right|_{(1,0)}$$

which yields $\left.\frac{dy}{dx}\right|_{(1,0)} = 0.$

Related Rates In implicit differentiation, we differentiate an equation involving x and y, with y treated as a function of x. However, in some applications where x and y are related by an equation, they are functions of a third variable, for example, time t. If we differentiate such an equation with respect to t, we get a relationship between the rates of change $\frac{dy}{dt}$ and $\frac{dx}{dt}$. These derivatives are called *related rates*.

Example The radius of a circle is increasing at the rate of 3 cm per second. Find the rate of change of the area inside the circle when the radius is 5 cm.

Explanation Both the area A and the radius r of the circle are functions of time t. It is give that $\frac{dr}{dt} = 3$. The question is to find $\frac{dA}{dt}$ when $r = 5$.

Solution The area A and the radius r of the circle are related by

$$A = \pi r^2.$$

Differentiating both sides of the equation with respect to time t, we get

$$\begin{aligned} \frac{d}{dt}A &= \frac{d}{dt}\pi r^2 \\ \frac{dA}{dt} &= \pi \cdot 2r \cdot \frac{dr}{dt} && \text{Constant Multiple Rule and Chain Rule \& Power Rule.} \\ \frac{dA}{dt} &= 2\pi r \cdot 3 && \text{Given that } \frac{dr}{dt} = 3 \end{aligned}$$

Thus at the instant where $r = 5$, we have

$$\frac{dA}{dt} = 2\pi \cdot 5 \cdot 3 = 30\pi.$$

That is, the area is increasing at the rate of $30\pi\,\text{cm}^2$ per second. □

Example A point is moving along the graph of $4x^2 + y^2 = 8$. When the point is at $(1,2)$, its x-coordinate is increasing at the rate of 3 units per second. How fast is the y-coordinate changing at that moment?

Explanation The question is to find $\frac{dy}{dt}$ when $(x,y) = (1,2)$, given that $\frac{dx}{dt} = 3$ at that instant.

Solution Differentiating both sides of the equation with respect to time t, we get

$$\begin{aligned} \frac{d}{dt}(4x^2 + y^2) &= \frac{d}{dt}8 \\ \frac{d}{dt}4x^2 + \frac{d}{dt}y^2 &= 0 && \text{Term by Term Differentiation and Derivative of Constant} \\ 4 \cdot 2x \cdot \frac{dx}{dt} + 2y \cdot \frac{dy}{dt} &= 0 && \text{Chain Rule \& Power Rule} \end{aligned}$$

Solving for $\frac{\mathrm{d}y}{\mathrm{d}t}$, we get

$$\frac{\mathrm{d}y}{\mathrm{d}t} = \frac{-8x\frac{\mathrm{d}x}{\mathrm{d}t}}{2y} = \frac{-4x}{y} \cdot \frac{\mathrm{d}x}{\mathrm{d}t}.$$

At the given moment, we have

$$\frac{\mathrm{d}y}{\mathrm{d}t} = \frac{-4 \cdot 1}{2} \cdot 3 = -6.$$

That is, the y-coordinate is decreasing at a rate of 6 units per second. □

Exercise 9.2

1. For each of the following equations, find $\frac{\mathrm{d}y}{\mathrm{d}x}$.

   (a) $xy^2 - x^2 + y = 0$ (b) $x^3 + y^3 - 6xy = 0$

   (c) $x^5 + 4xy^3 - 3y^5 = 1$ (d) $x \sin y + y^2 = 1$

   (e) $\sin(x + y) = y \cos x$ (f) $\mathrm{e}^y = x^2 + y^2$

   (g) $x^3 - y^2 = x \ln y$ (h) $x\,\mathrm{e}^y + y \sin x = \ln(x + y)$

2. Foe each of the following curves represented by the given equations, find the slope at the indicated point.

   (a) $2y^3 + y^2 - x = 0$, $(3, 1)$

   (b) $xy^2 - 3y^3 + 8 = 0$, $(4, 2)$

   (c) $y \sin x + 3 \cos y = 3 + \cos x$, $(\frac{\pi}{2}, 0)$

   (d) $\ln y = 2y^2 - x + 1$, $(3, 1)$

   (e) $\ln(x + \sin y) = x^2 + 2\mathrm{e}^y - 3$, $(1, 0)$

   (f) $x^3 + y^2 + x^3\mathrm{e}^{y^2} = 1$, $(0, 1)$

3. A point is moving on the graph of $xy = 24$. When the point is at $(4, 6)$, its x-coordinate is increasing at 5 units per second. How fast is the y-coordinate changing at that moment?

4. The radius of a spherical balloon is increasing at the rate of 5 cm per minute. How fast is the volume changing when the radius is 8 cm?

5. A 3 m ladder is placed against a wall. Suppose that the foot of the ladder is pulled along the ground at the rate of 1 m per second. How fast is the top end of the ladder sliding down the wall at the time when the foot is 2 m from the wall?

9.3 More Curve Sketching

Example Let $f(x) = x \ln x$.

(1) Find and classify the critical number(s) of f.

(2) Find the interval(s) on which f is increasing or decreasing, convex or concave.

(3) Sketch the graph of f.

Solution

(1) & (2) First we note that the domain of f is $(0, \infty)$.

Differentiating f, we get
$$\begin{aligned} f'(x) &= \frac{d}{dx}(x \ln x) \\ &= x \cdot \frac{d}{dx} \ln x + \ln x \cdot \frac{d}{dx} x \\ &= x \cdot \frac{1}{x} + \ln x \\ &= 1 + \ln x. \end{aligned}$$

Solving $f'(x) = 0$,
$$\begin{aligned} 1 + \ln x &= 0 \\ \ln x &= -1 \end{aligned}$$
we get the critical number of f: $x_1 = e^{-1}$.

	$(0, e^{-1})$	(e^{-1}, ∞)
f'	$-$	$+$

The function f is decreasing on $(0, e^{-1})$ and increasing on (e^{-1}, ∞).

Thus f has a local minimum at $x_1 = e^{-1}$.

Differentiating f', we get
$$\begin{aligned} f''(x) &= \frac{d}{dx}(1 + \ln x) \\ &= \frac{1}{x}. \end{aligned}$$

	$(0, \infty)$
f''	$+$

The function f is convex on $(0, \infty)$.

(3) The graph of f is shown in Figure 9.1.

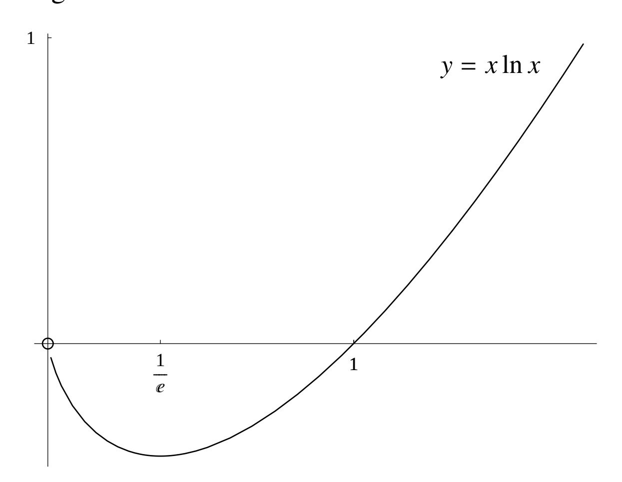


Figure 9.1

□

Example Sketch the graph of $f(x) = \sin x + \cos x$ for $0 \leq x \leq 2\pi$. On the graph, indicate the local extremum points and inflection points.

Solution Differentiating f, we get
$$\begin{aligned} f'(x) &= \frac{d}{dx}(\sin x + \cos x) \\ &= \cos x - \sin x. \end{aligned}$$

Solving $f'(x) = 0$,
$$\begin{aligned} \cos x &= \sin x \\ \tan x &= 1 \end{aligned}$$
we get the critical number of f in the interval $(0, 2\pi)$: $x_1 = \frac{\pi}{4}$ and $x_2 = \frac{5\pi}{4}$.

	$(0, \frac{\pi}{4})$	$(\frac{\pi}{4}, \frac{5\pi}{4})$	$(\frac{5\pi}{4}, 2\pi)$
f'	+	−	+

The function f is increasing on $(0, \frac{\pi}{4})$, decreasing on $(\frac{\pi}{4}, \frac{5\pi}{4})$ and increasing on $(\frac{5\pi}{4}, 2\pi)$.

Thus $(\frac{\pi}{4}, f(\frac{\pi}{4})) = (\frac{\pi}{4}, \sqrt{2})$ is a local maximum point and $(\frac{5\pi}{4}, f(\frac{5\pi}{4})) = (\frac{5\pi}{4}, -\sqrt{2})$ is a local minimum point of the graph.

Differentiating f', we get
$$\begin{aligned} f''(x) &= \frac{\mathrm{d}}{\mathrm{d}x}(\cos x - \sin x) \\ &= -\sin x - \cos x. \end{aligned}$$

Solving $f''(x) = 0$ in the interval $[0, 2\pi]$
$$\begin{aligned} -\cos x &= \sin x \\ -1 &= \tan x \end{aligned}$$
we get the zeros of f'' in the interval $(0, 2\pi)$: $x_3 = \frac{3\pi}{4}$ and $x_4 = \frac{7\pi}{4}$.

	$(0, \frac{3\pi}{4})$	$(\frac{3\pi}{4}, \frac{7\pi}{4})$	$(\frac{7\pi}{4}, 2\pi)$
f''	−	+	−

The function f is concave on $(0, \frac{3\pi}{4})$, convex on $(\frac{3\pi}{4}, \frac{7\pi}{4})$ and concave on $(\frac{7\pi}{4}, 2\pi)$.

Thus $(\frac{3\pi}{4}, f(\frac{3\pi}{4})) = (\frac{3\pi}{4}, 0)$ and $(\frac{7\pi}{4}, f(\frac{7\pi}{4})) = (\frac{7\pi}{4}, 0)$ are inflection points of the graph of f.

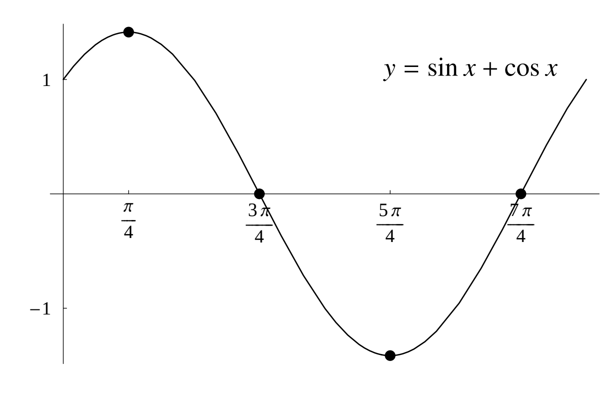


Figure 9.2

□

Remark

- Because the sine and cosine functions are periodic with period 2π, we can use the above graph to get the whole graph of f.

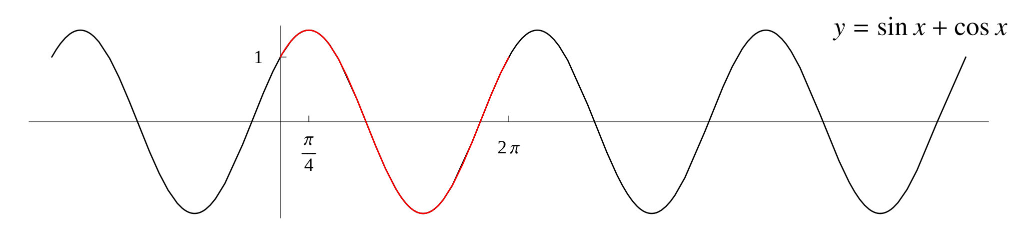


Figure 9.3

- Note that

$$\begin{aligned} f(x) &= \sin x + \cos x \\ &= \sqrt{2}\left(\frac{1}{\sqrt{2}}\sin x + \frac{1}{\sqrt{2}}\cos x\right) \\ &= \sqrt{2}\left(\cos\frac{\pi}{4}\sin x + \sin\frac{\pi}{4}\cos x\right) \\ &= \sqrt{2}\left(\sin\left(x + \frac{\pi}{4}\right)\right) \qquad \text{Compound angle formula} \end{aligned}$$

  Thus the graph of f can be obtained from that of the sine function by shifting it $\frac{\pi}{4}$ units to the left and then amplifying it by a factor of $\sqrt{2}$.

Exercise 9.3

1. For each of the following equations, sketch its graph.

   (a) $y = e^{-x^2}$ (b) $y = x\,e^{-x}$

   (c) $y = x - \ln x$ (d) $y = \dfrac{\ln x}{x}$

   (e) $y = x - 2\sin x$ (f) $y = \sin^2 x$

   (g) $y = x - \sqrt{1 - x^2}$ (h) $y = \dfrac{1}{\sqrt{1 + 2x^2}}$

9.4 More Extremum Problems

Example Among all line segments that stretch from points on the positive x-axis to points on the positive y-axis and passes through the point $(5, 2)$, find the one that has shortest length.

Explanation The length of the line segment can be expressed as a function of any one of the following: (1) the x-intercept of the line segment; (2) the slope of the line segment; (3) the angle between the line segment and the x-axis.

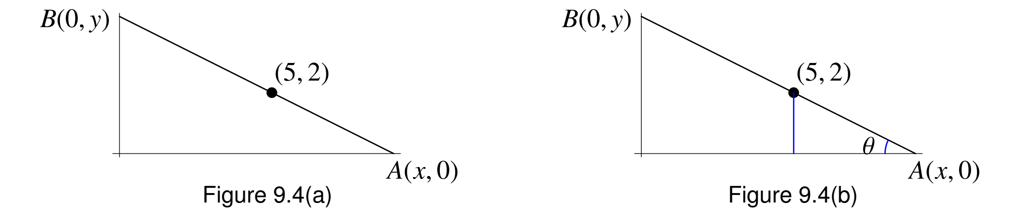


Figure 9.4(a) Figure 9.4(b)

Note that in Figure 9.4(a), we have $x > 5$ and in Figure 9.4(b), we have $0 < \theta < \frac{\pi}{2}$. Moreover, if m denotes the slope of the line segment, then we have $m < 0$.

Solution 1 Let $A(x, 0)$ and $B(0, y)$ be the points of intersection of the line segment with the x- and y-axes respectively. We want to minimize the length

$$L = \sqrt{x^2 + y^2}.$$

Since the line segment passes through $(5, 2)$, we get the following relationship between x and y.

$$\begin{aligned} \frac{y-2}{0-5} &= \frac{2-0}{5-x} \\ y &= \frac{-10}{5-x} + 2 \\ &= \frac{2x}{x-5} \end{aligned}$$

Therefore, we have $L = \sqrt{x^2 + \left(\dfrac{2x}{x-5}\right)^2}$.

Since L is a minimum when L^2 is a minimum, we consider minimizing

$$f(x) = x^2 + \frac{4x^2}{(x-5)^2}, \qquad x > 5.$$

Differentiating, we get

$$\begin{aligned} f'(x) &= 2x + \frac{(x-5)^2 \cdot \frac{\mathrm{d}}{\mathrm{d}x}(4x^2) - 4x^2 \cdot \frac{\mathrm{d}}{\mathrm{d}x}(x-5)^2}{((x-5)^2)^2} \\ &= 2x + \frac{(x-5)^2 \cdot 8x - 4x^2 \cdot 2(x-5)}{(x-5)^4} \\ &= 2x + \frac{8x(x-5) - 8x^2}{(x-5)^3} \\ &= 2x + \frac{-40x}{(x-5)^3}. \end{aligned}$$

Solving $f'(x) = 0$,

$$\begin{aligned} 2x &= \frac{40x}{(x-5)^3} \\ (x-5)^3 &= 20 \qquad \text{since } x > 5 \text{ implies } x \neq 0 \\ x - 5 &= 20^{\frac{1}{3}} \end{aligned}$$

we get the critical number of f in $(5, \infty)$: $x_1 = 5 + \sqrt[3]{20}$.

	$(5, 5 + \sqrt[3]{20})$	$(5 + \sqrt[3]{20}, \infty)$
f'	$-$	$+$

Note that $f'(x) = \dfrac{2x(x-5)^3 - 40x}{(x-5)^3} = \dfrac{2x[(x-5)^3 - 20]}{(x-5)^3}$.

Since f is decreasing on $(5, 5 + \sqrt[3]{20})$ and increasing on $(5 + \sqrt[3]{20}, \infty)$, it follows that f attains its absolute minimum at x_1. Therefore, the shortest line segment is the one that has x-intercept equal to $(5 + \sqrt[3]{20}, 0)$. □

Solution 2 Let $A(x, 0)$ and $B(0, y)$ be the points of intersection of the line segment with the x- and y-axes respectively. We want to minimize the length

$$L = \sqrt{x^2 + y^2}.$$

Note that both x and y are functions of the slope m of the line segment:

$$\begin{aligned} \frac{2-0}{5-x} &= m \\ \frac{2}{m} &= 5-x \\ x &= 5-\frac{2}{m} \end{aligned} \qquad \text{and} \qquad \begin{aligned} \frac{y-2}{0-5} &= m \\ y-2 &= -5m \\ y &= 2-5m \end{aligned}$$

Therefore, we have $L = \sqrt{\left(5-\frac{2}{m}\right)^2 + (2-5m)^2} = \sqrt{25m^2 - 20m + 29 - \frac{20}{m} + \frac{4}{m^2}}$

Since L is a minimum when L^2 is a minimum, we consider minimizing

$$f(m) = 25m^2 - 20m + 29 - \frac{20}{m} + \frac{4}{m^2}, \qquad m < 0.$$

Differentiating, we get

$$\begin{aligned} f'(m) &= \frac{\mathrm{d}}{\mathrm{d}m}\left(25m^2 - 20m + 29 - 20m^{-1} + 4m^{-2}\right) \\ &= 50m - 20 + 20m^{-2} - 8m^{-3} \end{aligned}$$

Solving $f'(m) = 0$,

$$\begin{aligned} 50m - 20 + 20m^{-2} - 8m^{-3} &= 0 \\ 25m^4 - 10m^3 + 10m - 4 &= 0 && \text{Multiply both sides by } \frac{m^3}{2} \\ (5m-2)(5m^3+2) &= 0 && \text{Factor Theorem: } L.S. = 0 \text{ when } m = \frac{2}{5} \\ 5m^3 + 2 &= 0 && \text{Since } m < 0 \\ m^3 &= -\frac{2}{5}, \end{aligned}$$

we get the critical number of f: $\quad m_1 = \sqrt[3]{-\frac{2}{5}}$.

Differentiating f', we get

$$\begin{aligned} f''(m) &= \frac{\mathrm{d}}{\mathrm{d}m}\left(50m - 20 + 20m^{-2} - 8m^{-3}\right) \\ &= 50 - 40m^{-3} + 24m^{-4}. \end{aligned}$$

Note that

$$f''(m_1) = 50 - 40 \cdot \left(-\frac{5}{2}\right) + 24 \cdot \left(\frac{5}{2}\right)^{\frac{4}{3}} > 0$$

and that m_1 is the only critical number of f in the open interval $(-\infty, 0)$, it follows from the Second Derivative Test (Special Version) that f attains its global minimum at m_1. Therefore, the shortest line segment is the one that has slope equal to $\sqrt[3]{-\frac{2}{5}}$. □

Remark Corresponding to m_1, we have $x_1 = 5 - \dfrac{2}{\sqrt[3]{-\frac{2}{5}}} = 5 + 2 \cdot \left(\frac{5}{2}\right)^{\frac{1}{3}} = 5 + 20^{\frac{1}{3}}$.

Solution 3 Let θ be the angle between the line segment and the x-axis. By considering the two right-angled triangles shown in Figure 9.4(b), we have

$$\cos\theta = \frac{x}{L} \qquad \text{and} \qquad \tan\theta = \frac{2}{x-5},$$

from which we get

$$L = \frac{x}{\cos\theta} = \frac{5 + \frac{2}{\tan\theta}}{\cos\theta}.$$

We want to minimize

$$L(\theta) = \frac{5}{\cos\theta} + \frac{2}{\sin\theta}, \qquad 0 < \theta < \frac{\pi}{2}.$$

Differentiating, we get

$$\begin{aligned}
\frac{dL}{d\theta} &= \frac{d}{d\theta}\left(5 \cdot (\cos\theta)^{-1} + 2 \cdot (\sin\theta)^{-1}\right) \\
&= 5 \cdot (-1) \cdot (\cos\theta)^{-2} \cdot \frac{d}{d\theta}\cos\theta + 2 \cdot (-1) \cdot (\sin\theta)^{-2} \cdot \frac{d}{d\theta}\sin\theta \\
&= 5 \cdot (-1) \cdot (\cos\theta)^{-2} \cdot (-\sin\theta) + 2 \cdot (-1) \cdot (\sin\theta)^{-2} \cdot \cos\theta \\
&= \frac{5\sin\theta}{\cos^2\theta} - \frac{2\cos\theta}{\sin^2\theta}.
\end{aligned}$$

Solving $\frac{dL}{d\theta} = 0$,

$$\begin{aligned}
\frac{5\sin\theta}{\cos^2\theta} &= \frac{2\cos\theta}{\sin^2\theta} \\
5\sin^3\theta &= 2\cos^3\theta \\
\tan^3\theta &= \frac{2}{5},
\end{aligned}$$

we get the critical number of L in $(0, \frac{\pi}{2})$: $\theta_1 = \tan^{-1}\sqrt[3]{\frac{2}{5}}$.

	$(0, \tan^{-1}\sqrt[3]{\frac{2}{5}})$	$(\tan^{-1}\sqrt[3]{\frac{2}{5}}, \frac{\pi}{2})$
$\frac{dL}{d\theta}$	$-$	$+$

Note that $\frac{dL}{d\theta} = \frac{5\sin^3\theta - 2\cos^2\theta}{\sin^2\theta\cos^2\theta} = \frac{5\cos^3\theta\,(\tan^3\theta - \frac{2}{5})}{\sin^2\theta\cos^2\theta}$.

Since L is decreasing on $(0, \tan^{-1}\sqrt[3]{\frac{2}{5}})$ and increasing on $(\tan^{-1}\sqrt[3]{\frac{2}{5}}, \frac{\pi}{2})$, it follows that L attains its minimum at θ_1. Therefore, the shortest line segment is the one that makes an angle $\tan^{-1}\sqrt[3]{\frac{2}{5}}$ with the x-axis. □

Remark Corresponding to θ_1, we have $x_1 = 5 + \frac{2}{\sqrt[3]{\frac{2}{5}}} = 5 + \sqrt[3]{20}$.

Example A recording company has produced a new CD. Before launching a sales campaign, the marketing research department wants to determine the length of the campaign that will maximize total profits. From empirical data, it is estimated that the proportion of a target group of 50000 persons buying the CD after t days of TV promotion is given by $1 - e^{-0.06t}$. If \$20 is received for each CD sold and the promotion cost is $C(t) = 200000 + 12000t$.

(1) How many days of TV promotion should be used to maximize the profit?

(2) What is the maximum profit?

(3) What percentage of the target group will have purchased the CD when the maximum profit is reached?

Explanation The number of days is a positive integer. In order to apply differentiation, we enlarge the domain to $(0, \infty)$.

Solution The revenue (in dollars) after t days of promotion is

$$R(t) = 20 \times 50000 \times (1 - e^{-0.06t}).$$

Therefore, the profit (in dollars) is

$$P(t) = 1000000(1 - e^{-0.06t}) - 200000 - 12000t.$$

We want to maximize $P(t)$ for positive integers t. First, we consider P as a function with domain $(0, \infty)$. Differentiating, we get

$$\begin{aligned} P'(t) &= \frac{d}{dt}\Big(1000000(1 - e^{-0.06t}) - 200000 - 12000t\Big) \\ &= 1000000 \cdot (-e^{-0.06t}) \cdot \frac{d}{dt}(-0.06t) - 12000 \\ &= 60000e^{-0.06t} - 12000. \end{aligned}$$

Solving $P'(t) = 0$,

$$\begin{aligned} 60000e^{-0.06t} &= 12000 \\ e^{-0.06t} &= 0.2 \\ -0.06t &= \ln 0.2, \end{aligned}$$

we get the critical number of P in $(0, \infty)$: $t_1 = \frac{\ln 0.2}{-0.06}$.

	$(0, \frac{\ln 0.2}{-0.06})$	$(\frac{\ln 0.2}{-0.06}, \infty)$
P'	+	−

Since P is increasing on $(0, \frac{\ln 0.2}{-0.06})$ and decreasing on $(\frac{\ln 0.2}{-0.06}, \infty)$, it follows that on $(0, \infty)$, P attains its maximum at t_1. However $t_1 \approx 26.8$ is not an integer. Comparing the profit at $t_2 = 26$ and $t_3 = 27$:

t	26	27
P	277864	278101

we see that

(1) 27 days of TV promotion should be used to maximize the profit;

(2) the maximum profit is $278101;

(3) $1 - e^{-0.06 \times 27} \approx 80\%$ of the target group will have purchased the CD.

□

Exercise 9.4

1. Find the area of the largest rectangle that has one side on the x-axis and two vertices on the curve $y = e^{-x^2}$.
2. Suppose the price-demand equation for a product is determined from empirical data to be $p = 100e^{-0.05q}$ where q is the number of units sold. Find the production level and price that maximize revenue. What is the maximum revenue?
3. A lake polluted by bacteria is treated with an antibacterial chemical. After t days, the number N of bacteria per ml of water is approximated by

$$N(t) = 20(\frac{t}{12} - \ln(\frac{t}{12})) + 30$$

for $1 \le t \le 15$. During this time

(a) when will the number of bacteria be a minimum? what is the minimum?

(b) when will the number of bacteria be a maximum? what is the maximum?

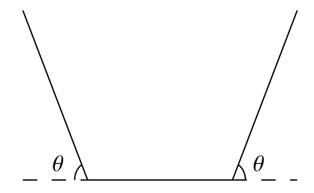
4. A company wishes to run an utility cable from a point A on the shore to an installation at point B on the island. The island is 6 km from the shore where C is the nearest point. Assume that the cable starts at a point A on the shore and runs along the shoreline, then angles and runs underwater to the island. It costs \$3200 per km to run the cable on land and \$4000 per km underwater. Find the point at which the line should begin to angle in order to yield the minimum cost if

   (a) the distance between A and C is 9 km;

   (b) the distance between A and C is 7 km.

5. A light source is to be placed directly above the center of a circular table of radius 1.5 m. The illumination at any point on the table is directly proportional to the sine of the angle between the table and the line joining the source and the point and inversely proportional to the square of the distance from the source. Find the height above the circle at which illumination on the edge of the table is maximized.

6. A long piece of metal one meter wide is to be bent in two places, $\frac{1}{3}$ meter from the two ends, to form a spillway so that its cross-section is an isosceles trapezoid. Find the angle θ at which the bend should be formed in order to obtain maximum possible flow along the spillway.



Chapter 10

More Integration

10.1 More Formulas

Using formulas for differentiation discussed in previous chapters, we get the corresponding formulas for integration.

Integration Formula 1 $\displaystyle\int x^r \,\mathrm{d}x = \frac{x^{r+1}}{r+1} + C \qquad \text{where } -1 \neq r \in \mathbb{R}$

Integration Formula 2 $\displaystyle\int \sin x \,\mathrm{d}x = -\cos x + C$

Integration Formula 3 $\displaystyle\int \cos x \,\mathrm{d}x = \sin x + C$

Integration Formula 4 $\displaystyle\int \sec^2 x \,\mathrm{d}x = \tan x + C$

Integration Formula 5 $\displaystyle\int \frac{1}{1+x^2} \,\mathrm{d}x = \tan^{-1} x + C$

Integration Formula 6 $\displaystyle\int \mathrm{e}^x \,\mathrm{d}x = \mathrm{e}^x + C$

Integration Formula 7 $\displaystyle\int \frac{1}{x} \,\mathrm{d}x = \ln|x| + C$

Explanation For each of the above formulas, the equality is valid on every open interval on which the integrand is defined. For example, Formula 4 means that on every open interval not containing any real number in the form $\frac{k\pi}{2}$ where k is an odd integer, the function $\tan x$ is an antiderivative of the function $\sec^2 x$.

The formulas can be proved directly by differentiating the functions on the right side. Below we give the proofs for (1), (2) and (7). For (7), since the domain of the function x^{-1} is $\mathbb{R} \setminus \{0\}$, we have to consider two cases: $x > 0$ and $x < 0$.

Proof for (1) Let r be a constant different from -1. On every open interval in which the function x^{r+1} is defined (hence the function x^r is also defined), we have

$$\begin{aligned}
\frac{\mathrm{d}}{\mathrm{d}x}\frac{x^{r+1}}{r+1} &= \frac{1}{r+1}\cdot\frac{\mathrm{d}}{\mathrm{d}x}x^{r+1} && \text{Constant Multiple Rule for Differentiation}\\
&= \frac{1}{r+1}\cdot(r+1)\cdot x^{r+1-1} && \text{(General) Power Rule for Differentiation}\\
&= x^r
\end{aligned}$$

Proof for (2) On $\mathbb{R}$, we have

$$\begin{aligned}
\frac{\mathrm{d}}{\mathrm{d}x}(-\cos x) &= (-1)\cdot\frac{\mathrm{d}}{\mathrm{d}x}\cos x && \text{Constant Multiple Rule for Differentiation}\\
&= (-1)\cdot(-\sin x) && \text{Derivative of cos}\\
&= \sin x
\end{aligned}$$

Proof for (7) To prove the result, we consider the following two cases:

(Case $x > 0$)

$$\begin{aligned}
\frac{\mathrm{d}}{\mathrm{d}x}\ln|x| &= \frac{\mathrm{d}}{\mathrm{d}x}\ln x && \text{Definition of } |x|\\
&= \frac{1}{x} && \text{Derivative of ln}
\end{aligned}$$

(Case $x < 0$)

$$\begin{aligned}
\frac{\mathrm{d}}{\mathrm{d}x}\ln|x| &= \frac{\mathrm{d}}{\mathrm{d}x}\ln(-x) && \text{Definition of } |x|\\
&= \frac{1}{-x}\cdot\frac{\mathrm{d}}{\mathrm{d}x}(-x) && \text{Chain Rule \& Derivative of ln}\\
&= \frac{1}{-x}\cdot(-1) && \text{Constant Multiple Rule for Differentiation and Power Rule for Differentiation}\\
&= \frac{1}{x}
\end{aligned}$$

□

Example Perform the following integration:

(1) $\int (x^2 + \sin x)\,\mathrm{d}x$

(2) $\int \left(1 - \frac{1}{x}\right)\mathrm{d}x$

(3) $\int (2\cos x + 3\mathrm{e}^x)\,\mathrm{d}x$

Solution

(1)
$$\begin{aligned}
\int (x^2 + \sin x)\,\mathrm{d}x &= \int x^2\,\mathrm{d}x + \int \sin x\,\mathrm{d}x && \text{Term by Term Integration}\\
&= \frac{x^3}{3} - \cos x + C && \text{Integration Formulas (1) \& (2)}
\end{aligned}$$

(2) $$\int \left(1 - \frac{1}{x}\right) \mathrm{d}x = \int 1 \, \mathrm{d}x - \int \frac{1}{x} \, \mathrm{d}x$$ Term by Term Integration

$$= x - \ln|x| + C$$ Integration Formulas (1) & (7)

(3) $$\int (2\cos x + 3\mathrm{e}^x) \, \mathrm{d}x = \int 2\cos x \, \mathrm{d}x + \int 3\mathrm{e}^x \, \mathrm{d}x$$ Term by Term Integration

$$= 2\int \cos x \, \mathrm{d}x + 3\int \mathrm{e}^x \, \mathrm{d}x$$ Constant Multiple for Integration

$$= 2\sin x + 3\mathrm{e}^x + C$$ Integration Formulas (3) & (6)

□

Example Evaluate the following definite integrals:

(1) $\displaystyle\int_0^{\frac{\pi}{2}} 3\sin x \, \mathrm{d}x$

(2) $\displaystyle\int_1^2 \left(\mathrm{e}^x + \frac{1}{x}\right) \mathrm{d}x$

Explanation In the first step of the solution, we apply rules and formulas for integration to find a primitive for the given integrand (on the closed interval determined by the limits of integration) together with the Fundamental Theorem of Calculus (Version 2).

Solution

(1) $$\int_0^{\frac{\pi}{2}} 3\sin x \, \mathrm{d}x = \Big[-3\cos x\Big]_0^{\frac{\pi}{2}}$$ Constant Multiple Rule, Integration Formula (2) & Fundamental Theorem of Calculus

$$= -3\cos\frac{\pi}{2} - (-3\cos 0)$$

$$= 3$$

(2) $$\int_1^2 \left(\mathrm{e}^x + \frac{1}{x}\right) \mathrm{d}x = \Big[\mathrm{e}^x + \ln|x|\Big]_1^2$$ Integration Formulas (6) & (7) & Fundamental Theorem of Calculus

$$= (\mathrm{e}^2 + \ln 2) - (\mathrm{e} + \ln 1)$$

$$= \mathrm{e}^2 - \mathrm{e} + \ln 2$$

□

Example Find the area of the (combined) region that lies between the x-axis and the graph of $y = \mathrm{e}^x - 1$ for $-1 \le x \le 2$.

Solution Note that for $-1 \le x \le 0$, the graph of $y = \mathrm{e}^x - 1$ is below the x-axis and for $0 \le x \le 2$, the graph is above the x-axis.

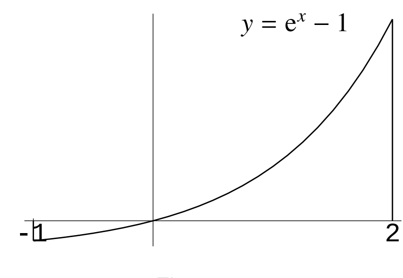


Figure 10.1

The required area A is

$$\begin{aligned}
A &= \int_{-1}^{0} [0 - (e^x - 1)]\,dx + \int_{0}^{2} [(e^x - 1) - 0]\,dx \\
&= \int_{-1}^{0} (1 - e^x)\,dx + \int_{0}^{2} (e^x - 1)\,dx \\
&= \Big[x - e^x\Big]_{-1}^{0} + \Big[e^x - x\Big]_{0}^{2} \\
&= [(-1 + 0) - (-e^{-1} - 1)] + [(e^2 - 2) - (1 - 0)] \\
&= e^2 + e^{-1} - 3
\end{aligned}$$

Term by Term Integration, Integration Formulas (1) & (6) & Fundamental Theorem of Calculus

□

Exercise 10.1

1. Perform the following integration:
   (a) $\int 3\sec^2 x\,dx$ (b) $\int (2e^x + \cos x)\,dx$
   (c) $\int \frac{2x+3}{x}\,dx$ (d) $\int (1 + \frac{1}{x})^2\,dx$
2. Evaluate the following definite integrals:
   (a) $\int_0^{\frac{\pi}{3}} 2\sin x\,dx$ (b) $\int_{-1}^{1} (2e^x + \sin x)\,dx$
   (c) $\int_{-4}^{-1} (e^x + \frac{1}{x})\,dx$ (d) $\int_1^2 \frac{2-x}{x}\,dx$
3. Find the area of the (combined) region that
   (a) lies between the x-axis and the graph of $y = \sin x$ for $0 \le x \le \pi$;
   (b) lies between the x-axis and the graph of $y = \frac{1}{x}$ for $\frac{1}{2} \le x \le e$;
   (c) lies under the graphs of $y = e^x$ and e^{-x} and above the x-axis for $-1 \le x \le 2$.

10.2 Substitution Method

Up to this stage, we can do simple integration using formulas and simple rules. For more complicated ones, like $\int x\,e^{x^2}\,dx$, we have to use some techniques for integration. In general, different forms of the integrand requires different techniques. In this section, we discuss a simple but important technique — the *substitution method*. It is the technique in integration that corresponds to the chain rule in differentiation.

Let $y = F(u)$ be a function of u and let $u = g(x)$ be a function of x. Then y can be considered as a function of x by taking the composition of F with g:

$$y = F\big(g(x)\big).$$

Suppose that the function g is differentiable on an open interval I and that the function F is differentiable on an open interval containing the image of I under g. Then by the Chain Rule, the composition function $F \circ g$ is differentiable on I and we have

$$\begin{aligned}
\frac{d}{dx}F(g(x)) &= \frac{dy}{dx} && \text{Definition} \\
&= \frac{dy}{du}\cdot\frac{du}{dx} && \text{Chain Rule} \\
&= F'(u)\cdot g'(x).
\end{aligned}$$

Writing everything in terms of x, we get

$$\frac{\mathrm{d}}{\mathrm{d}x}F(g(x)) = F'(g(x)) \cdot g'(x)$$

This is the chain rule expressed in an alternative way. Since integration is the reverse process of differentiation, we have (on the interval I)

$$\int F'(g(x)) \cdot g'(x)\,\mathrm{d}x = F(g(x)) + C.$$

Denoting $F' = f$, the above integration formula becomes

$$\int f(g(x))\,g'(x)\,\mathrm{d}x = F(g(x)) + C \tag{10.2.1}$$

Example Find $\int (x^2+1)^2 \cdot 2x\,\mathrm{d}x$.

Explanation By choosing f and g suitably, the integrand can be written as $f(g(x))\,g'(x)$. To apply (10.2.1), we can take any antiderivative for f.

Solution Put $f(x) = x^2$ and $g(x) = x^2 + 1$. Then we have

$$f(g(x)) = (x^2+1)^2 \quad \text{and} \quad g'(x) = 2x. \tag{10.2.2}$$

A primitive F for f is given by

$$F(x) = \frac{1}{3}x^3.$$

From these we get

$$\begin{aligned} \int (x^2+1)^2 \cdot 2x\,\mathrm{d}x &= \int f(g(x))\,g'(x)\,\mathrm{d}x && \text{By (10.2.2)} \\ &= F(g(x)) + C && \text{By (10.2.1)} \\ &= F(x^2+1) + C && \text{Definition of } g \\ &= \frac{1}{3}(x^2+1)^3 + C && \text{Definition of } F \end{aligned}$$

□

Remark In the given integral, the integrand is deliberately written as $(x^2+1)^2 \cdot 2x$. Usually, the integral is written as $\int 2x\,(x^2+1)^2\,\mathrm{d}x$.

In order to use (10.2.1), we have to choose *two* functions f and g suitably. Below we describe a more convenient way: change of variable (or substitution) — we only need to choose a suitable function g.

In (10.2.1), putting $u = g(x)$ and using $\mathrm{d}u = g'(x)\,\mathrm{d}x$ (see the explanation below), we get

$$\begin{aligned} \int f(g(x))\,g'(x)\,\mathrm{d}x &= \int f(u)\,\mathrm{d}u \\ &= F(u) + C. \end{aligned}$$

Explanation The notations $\mathrm{d}u$ and $\mathrm{d}x$ are called *differentials*. They are related by the fact that if Δx is small, then $\frac{\Delta u}{\Delta x}$ is approximately equal to $g'(x)$, that is,

$$\Delta u \approx g'(x)\Delta x.$$

In the limiting situation, we have $\mathrm{d}u = g'(x)\,\mathrm{d}x$.

Alternative procedure for the above example Put $u = x^2 + 1$. Then we have $\frac{\mathrm{d}u}{\mathrm{d}x} = 2x$ from which we get

$$\mathrm{d}u = 2x\,\mathrm{d}x. \tag{10.2.3}$$

Therefore, we have

$$\begin{aligned}
\int (x^2+1)^2 \cdot 2x\,\mathrm{d}x &= \int u^2\,\mathrm{d}u && \text{Substitution and (10.2.3)} \\
&= \frac{u^3}{3} + C && \text{Integration Formula (1)} \\
&= \frac{(x^2+1)^3}{3} + C && \text{Back substitution}
\end{aligned}$$

□

FAQ Do we get the same answer if we expand the integrand first?

Answer If we expand the integrand and then integrate term by term, we get

$$\begin{aligned}
\int (x^2+1)^2 \cdot 2x\,\mathrm{d}x &= \int 2x(x^4 + 2x^2 + 1)\,\mathrm{d}x \\
&= \int (2x^5 + 4x^3 + 2x)\,\mathrm{d}x \\
&= \frac{1}{3}x^6 + x^4 + x^2 + C.
\end{aligned}$$

The result obtained by the substitution method is

$$\begin{aligned}
\frac{1}{3}(x^2+1)^3 + C &= \frac{1}{3}(x^6 + 3x^4 + 3x^2 + 1) + C \\
&= \frac{1}{3}x^6 + x^4 + x^2 + \frac{1}{3} + C.
\end{aligned}$$

Although these two answers "look different", they represent the same family of functions. □

Remark If we change the integration to be

$$\int \sqrt{x^2+1} \cdot 2x\,\mathrm{d}x,$$

we can still use the substitution method but not the method by expansion.

Steps for the Substitution Method

(1) Define a new variable $u = g(x)$, where $g(x)$ is chosen in such a way that $g'(x)$ "is a factor" of the integrand and that when written in terms of u, the integrand is simpler than when written in terms of x.

(2) Transform the integral with respect to x into an integral with respect to u by replacing $g(x)$ everywhere by u and $g'(x)\,\mathrm{d}x$ by $\mathrm{d}u$.

(3) Integrate the resulting function of u.

(4) Substitute back $u = g(x)$ to express the result in terms of x.

Example Find $\int \frac{\ln x}{x}\,\mathrm{d}x$.

Explanation After choosing $u = g(x)$, we have $\frac{\mathrm{d}u}{\mathrm{d}x} = g'(x)$ which yields $\mathrm{d}u = g'(x)\,\mathrm{d}x$. Usually, the intermediate step is omitted. In the solution below, in the first equality in the equation array, we just rewrite the integrand so that substitution can be applied. Usually, this step is done in the head.

Solution Put $u = \ln x$. Then we have $\mathrm{d}u = \frac{1}{x}\,\mathrm{d}x$. From these we get

$$\begin{aligned}
\int \frac{\ln x}{x}\,\mathrm{d}x &= \int \ln x \cdot \frac{1}{x}\,\mathrm{d}x \\
&= \int u\,\mathrm{d}u && \text{Substitution} \\
&= \frac{u^2}{2} + C && \text{Integration Formula (1)} \\
&= \frac{1}{2}(\ln x)^2 + C && \text{Back substitution}
\end{aligned}$$

□

Remark Instead of writing down the substitution u explicitly, some authors use the following alternative steps:

$$\begin{aligned}
\int \frac{\ln x}{x}\,\mathrm{d}x &= \int \ln x\,\mathrm{d}(\ln x) \\
&= \frac{(\ln x)^2}{2} + C.
\end{aligned}$$

Example Find $\int x^2 \mathrm{e}^{x^3}\,\mathrm{d}x$.

Explanation If we choose $g(x) = x^3$. Then we have $g'(x) = 3x^2$. Although the factor 3 doesn't appear in the integrand, we can create it by writing $1 = 3 \cdot \frac{1}{3}$.

Solution Put $u = x^3$. Then we have $\mathrm{d}u = 3x^2\,\mathrm{d}x$. From these we get

$$\begin{aligned}
\int x^2 \mathrm{e}^{x^3}\,\mathrm{d}x &= \int \frac{1}{3}\mathrm{e}^{x^3} \cdot 3x^2\,\mathrm{d}x \\
&= \int \frac{1}{3}\mathrm{e}^{u}\,\mathrm{d}u && \text{Substitution} \\
&= \frac{1}{3}\mathrm{e}^{u} + C && \text{Constant Multiple Rule and Integration Formula (6)} \\
&= \frac{1}{3}\mathrm{e}^{x^3} + C && \text{Back substitution}
\end{aligned}$$

□

Example Find $\int \sin(2x - 3)\,\mathrm{d}x$.

Explanation In order to apply the substitution method, the integrands should be a product of two factors (see (10.2.1). Note that $\sin(2x - 3)$ can be written as $\sin(2x - 3) \cdot 1$. Moreover, the derivative of $(2x - 3)$ is 2 which is a multiple of 1.

Solution Put $u = 2x - 3$. Then we have $\mathrm{d}u = 2\,\mathrm{d}x$. From these we get

$$\begin{aligned}
\int \sin(2x-3)\,\mathrm{d}x &= \int \frac{1}{2}\sin(2x-3)\cdot 2\,\mathrm{d}x \\
&= \int \frac{1}{2}\sin u\,\mathrm{d}u && \text{Substitution} \\
&= \frac{1}{2}\cdot(-\cos u) && \text{Constant Multiple Rule and Integration Formula (2)} \\
&= -\frac{1}{2}\cos(2x-3) + C && \text{Back substitution}
\end{aligned}$$

□

The following rule can be obtained using the method for the above example.

Linear Change of Variable Rule Suppose that $\int f(x)\,\mathrm{d}x = F(x) + C$, $\alpha < x < \beta$. Then for every constants a and b with $a \neq 0$, we have

$$\int f(ax+b)\,\mathrm{d}x = \frac{1}{a}F(ax+b) + C, \quad \frac{\alpha - b}{a} < x < \frac{\beta - b}{a}$$

Proof The given equality means that $F'(x) = f(x)$ for all $x \in (\alpha, \beta)$. From this we get

$$\begin{aligned}
\frac{\mathrm{d}}{\mathrm{d}x}\left(\frac{1}{a}F(ax+b)\right) &= \frac{1}{a}\cdot\frac{\mathrm{d}}{\mathrm{d}x}F(ax+b) && \text{Constant Multiple Rule} \\
&= \frac{1}{a}\cdot F'(ax+b)\cdot\frac{\mathrm{d}}{\mathrm{d}x}(ax+b) && \text{Chain Rule} \\
&= \frac{1}{a}\cdot F'(ax+b)\cdot a && \text{Derivative of Polynomial} \\
&= F'(ax+b) \\
&= f(ax+b) \qquad \text{for } \alpha < ax+b < \beta.
\end{aligned}$$

Note that $\alpha < ax + b < \beta$ is the same as $\frac{\alpha - b}{a} < x < \frac{\beta - b}{a}$. Hence the required result follows. □

In the substitution method, most authors use u to be the new variable. Thus the method is usually called u-substitution.

A Guide for u-substitution Treat the integrand as a product of two functions of x. Choose u to be an expression appearing in one of the two functions such that $\frac{\mathrm{d}u}{\mathrm{d}x}$ is the other function or a multiple of the other function. If such an expression can be found, then the integral can be written as $\int f(u)\,\mathrm{d}u$ using substitution.

Remark The examples given in this section are chosen so that suitable u-substitutions can be used. If we change the integrands slightly, there may not be any suitable u-substitution. For example, we can use u-substitution to find $\int x\,\mathrm{e}^{x^2}\,\mathrm{d}x$. However, if we change the integral to be $\int x\,\mathrm{e}^{x}\,\mathrm{d}x$, we can't use u-substitution. Instead, we can use a technique called *Integration by Parts*. It is the technique in integration that corresponds to the Product Rule in differentiation. A brief introduction to this technique will be given in a later section.

Integration is difficult. In fact, there are functions that *can't be integrated*. For example, we can't express $\int \mathrm{e}^{x^2}\,\mathrm{d}x$ using functions that we have discussed.

Substitution Method for Definite Integrals

To find definite integrals using u-substitution, one method is to find antiderivatives for the integrands and then apply the Fundamental Theorem of Calculus. Alternatively, we may change of the definite integrals to ones in terms of u by changing the limits of integration accordingly:

$$\int_a^b f(g(x))\, g'(x)\, \mathrm{d}x = \int_{g(a)}^{g(b)} f(u)\, \mathrm{d}u \tag{10.2.4}$$

where g is a continuous function on $[a, b]$ and f is a function defined and continuous on an open interval I containing the image of $[a, b]$ under g.

Proof Let F be a function such that $F' = f$ on I. Note that $F \circ g$ is a primitive for $(f \circ g) \cdot g'$ on $[a, b]$. Thus we have

$$\begin{aligned}
\int_a^b f(g(x))\, g'(x)\, \mathrm{d}x &= \Big[F(g(x))\Big]_a^b && \text{Fundamental Theorem of Calculus} \\
&= F(g(b)) - F(g(a)) \\
&= \Big[F(u)\Big]_{g(a)}^{g(b)} \\
&= \int_{g(a)}^{g(b)} f(u)\, \mathrm{d}u && \text{Fundamental Theorem of Calculus}
\end{aligned}$$

□

Example Evaluate $\displaystyle\int_0^4 x\sqrt{x^2+9}\, \mathrm{d}x$.

Solution 1 Put $u = x^2 + 9$. Then we have $\mathrm{d}u = 2x\, \mathrm{d}x$. From these we get

$$\begin{aligned}
\int x\sqrt{x^2+9}\, \mathrm{d}x &= \int \frac{1}{2}\sqrt{x^2+9} \cdot 2x\, \mathrm{d}x \\
&= \frac{1}{2}\int u^{\frac{1}{2}}\, \mathrm{d}u \\
&= \frac{1}{2} \cdot \frac{u^{\frac{3}{2}}}{\frac{3}{2}} + C \\
&= \frac{1}{3}(x^2+9)^{\frac{3}{2}} + C.
\end{aligned}$$

By the Fundamental Theorem of Calculus, we have

$$\begin{aligned}
\int_0^4 x\sqrt{x^2+9}\, \mathrm{d}x &= \left[\frac{1}{3}(x^2+9)^{\frac{3}{2}}\right]_0^4 \\
&= \frac{1}{3} \cdot 125 - \frac{1}{3} \cdot 27 \\
&= \frac{98}{3}
\end{aligned}$$

Solution 2 Put $u = x^2 + 9$. Then we have $\mathrm{d}u = 2x\, \mathrm{d}x$.

Note that when $x = 4$, $u = 25$ and when $x = 0$, $u = 9$.

Therefore by (10.2.4), we have

$$\begin{aligned}
\int_0^4 x\sqrt{x^2+9}\,\mathrm{d}x &= \int_0^4 \frac{1}{2}\sqrt{x^2+9}\cdot 2x\,\mathrm{d}x \\
&= \frac{1}{2}\int_9^{25} u^{\frac{1}{2}}\,\mathrm{d}u \\
&= \frac{1}{2}\left[\frac{u^{\frac{3}{2}}}{\frac{3}{2}}\right]_9^{25} \\
&= \frac{1}{2}\left(\frac{2}{3}\cdot 125 - \frac{2}{3}\cdot 27\right) \\
&= \frac{98}{3}
\end{aligned}$$

□

Remark Instead of writing down the substitution u explicitly, some authors use the following alternative steps:

$$\begin{aligned}
\int_0^4 x\sqrt{x^2+9}\,\mathrm{d}x &= \int_0^4 \sqrt{x^2+9}\,\frac{\mathrm{d}(x^2+9)}{2} \\
&= \frac{1}{2}\left[\frac{(x^2+9)^{\frac{3}{2}}}{\frac{3}{2}}\right]_0^4 \\
&\vdots
\end{aligned}$$

In the rest of this section, we will apply (10.2.4) to find definite integrals using u-substitution.

Example Evaluate $\displaystyle\int_0^1 (x+1)\mathrm{e}^{x^2+2x}\,\mathrm{d}x$.

Solution Put $u = x^2 + 2x$. Then we have $\mathrm{d}u = (2x+2)\,\mathrm{d}x$.

Note that when $x = 1$, $u = 3$ and when $x = 0$, $u = 0$.

Therefore we have

$$\begin{aligned}
\int_0^1 (x+1)\mathrm{e}^{x^2+2x}\,\mathrm{d}x &= \int_0^1 \frac{1}{2}\mathrm{e}^{x^2+2x}\cdot(2x+2)\,\mathrm{d}x \\
&= \int_0^3 \frac{1}{2}\mathrm{e}^u\,\mathrm{d}u \\
&= \left[\frac{1}{2}\mathrm{e}^u\right]_0^3 \\
&= \frac{1}{2}(\mathrm{e}^3 - 1)
\end{aligned}$$

□

Example Evaluate $\displaystyle\int_0^{\frac{\pi}{2}} \sin x\cos x\,\mathrm{d}x$.

Solution Put $u = \sin x$. Then we have $\mathrm{d}u = \cos x\,\mathrm{d}x$.

Note that when $x = \frac{\pi}{2}$, $u = 1$ and when $x = 0$, $u = 0$.

Therefore we have

$$\begin{aligned}\int_0^{\frac{\pi}{2}} \sin x \cos x \, \mathrm{d}x &= \int_0^1 u \, \mathrm{d}u \\ &= \left[\frac{u^2}{2}\right]_0^1 \\ &= \frac{1}{2}.\end{aligned}$$

□

Remark We can also use the u-substitution $u = \cos x$.

Example Find the area of the (combined) region that lies between the x-axis and the graph of $y = xe^{-x^2}$ for $-1 \le x \le 2$.

Solution Note that for $-1 \le x \le 0$, the graph of $y = xe^{-x^2}$ is below the x-axis and for $0 \le x \le 2$, the graph is above the x-axis.

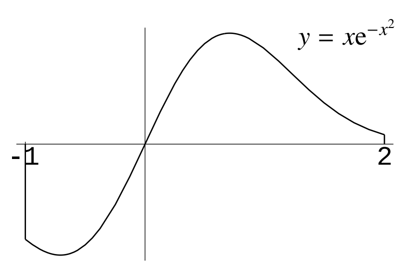


Figure 10.2

The required area A is

$$A = \int_{-1}^{0} (0 - xe^{-x^2}) \, \mathrm{d}x + \int_0^2 (xe^{-x^2} - 0) \, \mathrm{d}x.$$

Put $u = -x^2$. Then we have $\mathrm{d}u = -2x \, \mathrm{d}x$.

Note that when $x = -1$, $u = -1$; when $x = 0$, $u = 0$ and when $x = 2$, $u = -4$.

Therefore we have

$$\begin{aligned}A &= \int_{-1}^{0} \frac{1}{2} e^{-x^2} \cdot (-2x) \, \mathrm{d}x + \int_0^2 -\frac{1}{2} e^{-x^2} \cdot (-2x) \, \mathrm{d}x \\ &= \int_{-1}^{0} \frac{1}{2} e^u \, \mathrm{d}u + \int_0^{-4} -\frac{1}{2} e^u \, \mathrm{d}u \\ &= \left[\frac{1}{2} e^u\right]_{-1}^{0} + \left[-\frac{1}{2} e^u\right]_0^{-4} \\ &= \left(\frac{1}{2} - \frac{1}{2e}\right) + \left(-\frac{1}{2e^4} + \frac{1}{2}\right) \\ &= 1 - \frac{1}{2e} - \frac{1}{2e^4}\end{aligned}$$

□

Exercise 10.2

1. Perform the following integration:

(a) $\int 2x(x^2+1)^9 \, \mathrm{d}x$	(b) $\int x^4 \sqrt{x^5+6} \, \mathrm{d}x$	(c) $\int x \sin x^2 \, \mathrm{d}x$
(d) $\int \sin x \cos^2 x \, \mathrm{d}x$	(e) $\int 2xe^{x^2} \, \mathrm{d}x$	(f) $\int e^x \sec^2(e^x) \, \mathrm{d}x$
(g) $\int xe^{-x^2+1} \, \mathrm{d}x$	(h) $\int x^2 e^{x^3-1} \, \mathrm{d}x$	(i) $\int \frac{x}{x^2+1} \, \mathrm{d}x$
(j) $\int \frac{\sin \frac{1}{x}}{x^2} \, \mathrm{d}x$	(k) $\int (x+1)(x^2+2x+3)^7 \, \mathrm{d}x$	(l) $\int \frac{x^3+x}{(x^4+2x^2+3)^{11}} \, \mathrm{d}x$
(m) $\int (e^x - 3x)^4 (e^x - 3) \, \mathrm{d}x$	(n) $\int \frac{e^{\sqrt{x}}}{\sqrt{x}} \, \mathrm{d}x$	(o) $\int \frac{\ln(x+1)}{x+1} \, \mathrm{d}x$
(p) $\int \frac{1}{2x+7} \, \mathrm{d}x$	(q) $\int (x+1)^{15} \, \mathrm{d}x$	(r) $\int \frac{x}{\sqrt{x+1}} \, \mathrm{d}x$
(s) $\int x(x+1)^{15} \, \mathrm{d}x$	(t) $\int \frac{(x^2-1)e^{x+\frac{1}{x}}}{x^2} \, \mathrm{d}x$	

2. Evaluate the following definite integrals:

(a)	$\int_0^1 x(x^2+1)^5\,\mathrm{d}x$	(b)	$\int_0^1 x\mathrm{e}^{x^2+1}\,\mathrm{d}x$
(c)	$\int_0^1 x^2\cos x^3\,\mathrm{d}x$	(d)	$\int_0^2 \frac{1}{2x+3}\,\mathrm{d}x$
(e)	$\int_{-1}^0 \mathrm{e}^{x+1}\,\mathrm{d}x$	(f)	$\int_1^{\mathrm{e}^\pi} \frac{\sin(\ln x)}{x}\,\mathrm{d}x$
(g)	$\int_{-1}^1 (x-1)(x^2-2x)^4\,\mathrm{d}x$	(h)	$\int_{\mathrm{e}}^{\mathrm{e}^2} \frac{1}{x\ln x}\,\mathrm{d}x$
(i)	$\int_0^1 x(1-x)^7\,\mathrm{d}x$	(j)	$\int_0^8 \frac{x}{\sqrt{x+1}}\,\mathrm{d}x$

3. Find the area of the region that lies between
   (a) the x-axis and the graph of $y = x\sin x^2$ for $0 \le x \le \sqrt{\pi}$;
   (b) the graphs of $y = x$ and $y = x\mathrm{e}^{x^2}$ for $0 \le x \le 1$.

10.3 Integration of Rational Functions

Recall that rational functions are functions that can be written in the form $\frac{f(x)}{g(x)}$, where $f(x)$ and $g(x)$ are polynomials. If the degree of $f(x)$ is greater than or equal to that of $g(x)$, then by long division, we can find a polynomial $p(x)$ and a polynomial $r(x)$ with degree less than that of $g(x)$ such that

$$\frac{f(x)}{g(x)} = p(x) + \frac{r(x)}{g(x)} \qquad \text{for all } x \text{ with } g(x) \neq 0.$$

Since polynomial functions can be integrated easily, to integrate $\frac{f(x)}{g(x)}$, it suffices to know how to integrate $\frac{r(x)}{g(x)}$.

For the case where the degree of $g(x)$ is 1, the rational function $\frac{r(x)}{g(x)}$ takes the form $\frac{A}{ax+b}$ (where $a \neq 0$) which can be integrated as follows:

$$\begin{aligned} \int \frac{A}{ax+b}\,\mathrm{d}x &= A\int \frac{1}{ax+b}\,\mathrm{d}x && \text{Constant Multiple Rule} \\ &= A\cdot\frac{1}{a}\cdot\ln|ax+b| + C && \text{Integration Formula (7) \& Linear Change of Variable Rule} \end{aligned}$$

Below we discuss how to integrate rational functions where the degree of the denominator is 2 and the degree of the numerator is less than 2. Readers who want to know how to integrate rational functions where the degree of the denominator is greater than 2 may consult any (one-variable) calculus book.

To integrate rational functions in the form

$$\frac{Ax+B}{ax^2+bx+c}$$

where A, B, a, b, c are constants and $a \neq 0$, we consider the cases where the discriminant $b^2 - 4ac$ is positive, zero or negative:

(*Case 1*) $b^2 - 4ac > 0$

In this case, the denominator can be factorized as

$$ax^2 + bx + c = a(x - x_1)(x - x_2)$$

where x_1 and x_2 are the distinct real numbers. Moreover, there exists constants α and β such that

$$\frac{Ax+B}{ax^2+bx+c} = \frac{\alpha}{x-x_1} + \frac{\beta}{x-x_2} \tag{10.3.1}$$

for all $x \in \mathbb{R} \setminus \{x_1, x_2\}$. Note that the right-side can be integrated easily.

Terminology Fractions in the form

$$\frac{\alpha}{(x-x_1)^n},$$

where n is a positive integer, are called *partial fractions*. The sum in (10.3.1) is called the *partial-fraction decomposition* of the rational function on the left-side.

Example Find the partial-fraction decomposition of $\dfrac{x}{x^2-2x-3}$.

Solution Note that $x^2-2x-3 = (x-3)(x+1)$. The partial-fraction decomposition of the given rational function takes the form

$$\frac{x}{x^2-2x-3} = \frac{\alpha}{x-3} + \frac{\beta}{x+1}. \tag{10.3.2}$$

Multiplying both sides by $(x-3)(x+1)$, we get

$$x = \alpha(x+1) + \beta(x-3). \tag{10.3.3}$$

To find the constants α and β, we can use the *compare coefficient method* or the *substitution method*.

(*Compare Coefficient Method*) From (10.3.3), we get

$$x = (\alpha+\beta)x + (\alpha - 3\beta).$$

Comparing the coefficients of the x term and the constant term, we get

$$1 = \alpha + \beta \qquad \text{and} \qquad 0 = \alpha - 3\beta$$

respectively. Solving, we get $\alpha = \frac{3}{4}$ and $\beta = \frac{1}{4}$.

(*Substitution Method*) In (10.3.3), putting $x = -1$ and $x = 3$, we get

$$-1 = -4\beta \qquad \text{and} \qquad 3 = 4\alpha$$

respectively. Thus we have $\alpha = \frac{3}{4}$ and $\beta = \frac{1}{4}$.

Explanation α and β are constants such that (10.3.2) holds for all $x \in \mathbb{R} \setminus \{-1, 3\}$ which (by continuity of polynomial functions) implies that (10.3.3) holds for all $x \in \mathbb{R}$. To find α and β, we substitute $x = 3$ and $x = -1$ respectively. In fact, we can substitute any two values of x to get a system of two linear equations with knowns α and β.

Therefore, we have the following partial-fraction decomposition:

$$\frac{x}{x^2-2x-3} = \frac{\frac{3}{4}}{x-3} + \frac{\frac{1}{4}}{x+1}.$$

□

Example Find $\displaystyle\int \frac{x+1}{x^2-2x-3}\,\mathrm{d}x$.

Solution From the result of the preceding example, we get

$$\begin{aligned}\int \frac{x+1}{x^2-2x-3}\,\mathrm{d}x &= \int\left(\frac{3}{4}\cdot\frac{1}{x-3}+\frac{1}{4}\cdot\frac{1}{x+1}\right)\mathrm{d}x\\ &= \frac{3}{4}\int\frac{1}{x-3}\,\mathrm{d}x+\frac{1}{4}\int\frac{1}{x+1}\,\mathrm{d}x\\ &= \frac{3}{4}\ln|x-3|+\frac{1}{4}\ln|x+1|+C\end{aligned}$$

□

(*Case 2*) $b^2-4ac=0$

In this case, the denominator can be factorized as

$$ax^2+bx+c=a(x-x_1)^2$$

where x_1 is a real number. Moreover, there exists constants α and β such that

$$\frac{Ax+B}{ax^2+bx+c}=\frac{\alpha}{x-x_1}+\frac{\beta}{(x-x_1)^2} \tag{10.3.4}$$

for all $x\in\mathbb{R}\setminus\{x_1\}$. Note that the right-side can be integrated easily.

Terminology The sum in (10.3.4) is called the *partial-fraction decomposition* of the rational function on the left-side.

Example Find the partial-fraction decomposition of $\dfrac{2x+3}{x^2-2x+1}$.

Solution Note that $x^2-2x+1=(x-1)^2$. The partial-fraction decomposition of the given rational function takes the form

$$\frac{2x+3}{x^2-2x+1}=\frac{\alpha}{x-1}+\frac{\beta}{(x-1)^2}. \tag{10.3.5}$$

Multiplying both sides by $(x-1)^2$, we get

$$2x+3=\alpha(x-1)+\beta. \tag{10.3.6}$$

To find the constants α and β, we can use any one of the following two methods:

(*Compare Coefficient Method*) From (10.3.6), we get

$$2x+3=\alpha x+(\beta-\alpha).$$

Comparing the coefficients of the x term and the constant term, we get

$$2=\alpha \qquad \text{and} \qquad 3=\beta-\alpha$$

respectively, which yields $\beta=5$.

(*Substitution Method*) In (10.3.6), putting $x = 1$ and $x = 0$, we get

$$5 = \beta \qquad \text{and} \qquad 3 = -\alpha + \beta$$

respectively, which yields $\alpha = 2$.

Therefore, we have the following partial-fraction decomposition:

$$\frac{2x+3}{x^2-2x+1} = \frac{2}{x-1} + \frac{5}{(x-1)^2}.$$

□

Example Find $\displaystyle\int \frac{2x+3}{x^2-2x+1}\,\mathrm{d}x$.

Solution From the result of the preceding example, we get

$$\begin{aligned}
\int \frac{2x+3}{x^2-2x+1}\,\mathrm{d}x &= \int \left(\frac{2}{x-1} + \frac{5}{(x-1)^2}\right)\mathrm{d}x \\
&= 2\int \frac{1}{x-1}\,\mathrm{d}x + 5\int \frac{1}{(x-1)^2}\,\mathrm{d}x \\
&= 2\ln|x-1| + 5\int (x-1)^{-2}\,\mathrm{d}x \\
&= 2\ln|x-1| - 5(x-1)^{-1} + C
\end{aligned}$$

□

(*Case 3*) $b^2 - 4ac < 0$

In this case, the denominator can be written as

$$ax^2 + bx + c = a\Big((x+s)^2 + t^2\Big)$$

where s and t are real numbers and $t \neq 0$. Before discussing how to find

$$\int \frac{Ax+B}{ax^2+bx+c}\,\mathrm{d}x,$$

in general, we consider the special cases where $A = 0$ or where $Ax + B$ is a multiple of the derivative of $ax^2 + bx + c$.

(*Subcase 3a*) $A = 0$

In this case, we have

$$\begin{aligned}
\int \frac{Ax+B}{ax^2+bx+c}\,\mathrm{d}x &= \int \frac{B}{a((x+s)^2+t^2)}\,\mathrm{d}x \\
&= \frac{B}{a}\int \frac{1}{t^2\left[\left(\frac{1}{t}x+\frac{s}{t}\right)^2+1\right]}\,\mathrm{d}x \\
&= \frac{B}{a}\cdot\frac{1}{t^2}\cdot\frac{1}{\frac{1}{t}}\tan^{-1}\left(\tfrac{1}{t}x+\tfrac{s}{t}\right) + C && \text{Integration Formula 5 \& Linear Change of Variable Rule} \\
&= \frac{B}{at}\tan^{-1}\frac{x+s}{t} + C
\end{aligned}$$

Example Find $\displaystyle\int \frac{1}{x^2+4x+13}\, dx$.

Explanation In the solution below, instead of applying the formula obtained above, we use a suitable u-substitution. The idea is to choose u so that $(x+2)^2+9 = (3u)^2+3^2$ (note that $\frac{1}{(3u)^2+3^2} = \frac{1}{9}\cdot\frac{1}{u^2+1}$ can be integrated easily).

Solution Note that $x^2+4x+5 = (x+2)^2+9$. Thus we have

$$\begin{aligned}
\int \frac{1}{x^2+4x+13}\,\mathrm{d}x &= \int \frac{1}{(x+2)^2+9}\,\mathrm{d}x \\
&= \int \frac{1}{(3u)^2+3^2}\cdot 3\,\mathrm{d}u && \text{Put } x+2=3u. \text{ Thus } \mathrm{d}x = 3\,\mathrm{d}u. \\
&= \frac{1}{3}\int \frac{1}{u^2+1}\,\mathrm{d}u \\
&= \frac{1}{3}\cdot \tan^{-1} u + C \\
&= \frac{1}{3}\tan^{-1}\frac{x+2}{3} + C
\end{aligned}$$

□

(*Subcase 3b*) $Ax+B = k(2ax+b)$ for some constant k

In this case, we have

$$\int \frac{Ax+B}{ax^2+bx+c}\,\mathrm{d}x = \int \frac{k(2ax+b)}{ax^2+bx+c}\,\mathrm{d}x$$

which can be integrated using substitution $u = ax^2+bx+c$.

Example Find $\displaystyle\int \frac{x+1}{2x^2+4x+5}\,\mathrm{d}x$.

Solution Put $u = 2x^2+4x+5$. Then we have $\mathrm{d}u = (4x+4)\,\mathrm{d}x = 4(x+1)\,\mathrm{d}x$. From these we get

$$\begin{aligned}
\int \frac{x+1}{2x^2+4x+5}\,\mathrm{d}x &= \int \frac{1}{4}\cdot\frac{1}{2x^2+4x+5}\cdot 4(x+1)\,\mathrm{d}x \\
&= \frac{1}{4}\int \frac{1}{u}\,\mathrm{d}u \\
&= \frac{1}{4}\cdot \ln|u| + C \\
&= \frac{1}{4}\ln(2x^2+4x+5) + C
\end{aligned}$$

□

Remark In the last step, the absolute value sign is omitted. This is because $2x^2+4x+5$ is always positive.

(*Case 3 in general*) To integrate $\frac{Ax+B}{ax^2+bx+c}$ where $b^2-4ac < 0$, we rewrite the numerator as a sum of two terms—the first one is a multiple of the derivative of the denominator and the second one is a constant.

Example Find $\displaystyle\int \frac{2x+3}{x^2+4x+13}\,\mathrm{d}x$.

Solution Note that $\frac{d}{dx}(x^2 + 4x + 13) = 2x + 4$. Writing $2x + 3 = (2x + 4) - 1$, we have

$$\int \frac{2x+3}{x^2+4x+13}\,dx = \int \frac{2x+4}{x^2+4x+13}\,dx - \int \frac{1}{x^2+4x+13}\,dx.$$

For the first integral, we put $u = x^2 + 4x + 13$ which gives $du = (2x + 4)\,dx$ and so we have

$$\begin{aligned}\int \frac{2x+4}{x^2+4x+13}\,dx &= \int \frac{1}{u}\,du \\ &= \ln|u| + c \\ &= \ln(x^2+4x+13) + C.\end{aligned}$$

For the second integral, by a previous result, we have

$$\int \frac{1}{x^2+4x+13}\,dx = \frac{1}{3}\tan^{-1}\frac{x+2}{3} + C.$$

Combining the two results, we get

$$\int \frac{2x+3}{x^2+4x+13}\,dx = \ln(x^2+4x+13) - \frac{1}{3}\tan^{-1}\frac{x+2}{3} + C.$$

□

Remark If $Ax + B = k(2ax + b)$, the method discussed in Subcase 3b works also for the cases where $b^2 - 4ac$ is positive or zero.

Example Find $\int \frac{x-1}{x^2-2x-3}\,dx$.

Solution

(*Method 1*) Put $u = x^2 - 2x - 3$. Then we have $du = (2x - 2)\,dx = 2(x - 1)\,dx$. From these we get

$$\begin{aligned}\int \frac{x-1}{x^2-2x-3}\,dx &= \int \frac{1}{2}\cdot\frac{2(x-1)}{x^2-2x-3}\,dx \\ &= \frac{1}{2}\int \frac{1}{u}\,du \\ &= \frac{1}{2}\cdot\ln|u| + C \\ &= \frac{1}{2}\ln|x^2-2x-3| + C\end{aligned}$$

(*Method 2*) Note that $x^2 - 2x - 3 = (x - 3)(x + 1)$. The partial-fraction decomposition of the integrand takes the form

$$\frac{x-1}{x^2-2x-3} = \frac{\alpha}{x-3} + \frac{\beta}{x+1}.$$

Multiplying both sides by $(x - 3)(x + 1)$, we get

$$x - 1 = \alpha(x + 1) + \beta(x - 3).$$

Putting $x = -1$ and $x = 3$, we get

$$-2 = -4\beta \qquad \text{and} \qquad 2 = 4\alpha$$

respectively which yields $\alpha = \beta = \frac{1}{2}$. Therefore we have

$$\begin{aligned}\int \frac{x-1}{x^2-2x-3}\,\mathrm{d}x &= \int \left(\frac{1}{2}\cdot\frac{1}{x-3}+\frac{1}{2}\cdot\frac{1}{x+1}\right)\mathrm{d}x \\ &= \frac{1}{2}\Big(\ln|x-3|+\ln|x+1|\Big)+C\end{aligned}$$

□

Remark Since $\ln|x^2-2x-3| = \ln\left(|x-3|\cdot|x+1|\right) = \ln|x-3|+\ln|x-1|$, the answers obtained in the above two solutions are the same.

10.4 Integration by Parts

The technique in integration that corresponds to the product rule in differentiation is called *integration by parts*. In this section, we give a brief discussion on this technique. Readers who want to know how to apply this technique to more examples may consult (one variable) calculus books.

Let f and g be functions that are differentiable on an open interval (a, b). By the product rule, we have

$$\frac{\mathrm{d}}{\mathrm{d}x}\Big(f(x)g(x)\Big) = f'(x)g(x) + f(x)g'(x), \qquad a < x < b$$

which, written in terms of integration, becomes

$$\int f'(x)g(x)\,\mathrm{d}x + \int f(x)g'(x)\,\mathrm{d}x = f(x)g(x), \qquad a < x < b.$$

If one of the two integrals on the left side is easy to find, then we can find the other one. By symmetry, we may assume that the first integral is easy to find, then in this case, we can find the second integral by the following:

$$\int f(x)g'(x)\,\mathrm{d}x = f(x)g(x) - \int f'(x)g(x)\,\mathrm{d}x \tag{10.4.1}$$

Remark For simplicity, in the above formula, the interval under consideration is omitted.

Below, we give an example to illustrate how to apply (10.4.1).

Example Find $\int x\,\mathrm{e}^x\,\mathrm{d}x$.

Explanation In the calculation below, note that $\int f'(x)g(x)\,\mathrm{d}x = \int \mathrm{e}^x\,\mathrm{d}x$ is easy to find.

Solution Put $f(x) = x$ and put $g(x) = \mathrm{e}^x$. Then we have $f'(x) = 1$ and $g'(x) = \mathrm{e}^x$. From these we get

$$\begin{aligned}\int x\,\mathrm{e}^x\,\mathrm{d}x &= \int f(x)g'(x)\,\mathrm{d}x && \text{Substitution} \\ &= f(x)g(x) - \int f'(x)g(x)\,\mathrm{d}x && \text{By (10.4.1)} \\ &= x\,\mathrm{e}^x - \int \mathrm{e}^x\,\mathrm{d}x && \text{Back substitution} \\ &= x\,\mathrm{e}^x - \mathrm{e}^x + C\end{aligned}$$

□

Remark There are infinitely many way to choose $g(x)$; we can add any constant to e^x. If we take $g(x) = \mathrm{e}^x + 1$, then we get

$$\begin{aligned}
\int x\,\mathrm{e}^x\,\mathrm{d}x &= \int f(x)g'(x)\,\mathrm{d}x \\
&= f(x)g(x) - \int f'(x)g(x)\,\mathrm{d}x \\
&= x\,(\mathrm{e}^x + 1) - \int (\mathrm{e}^x + 1)\,\mathrm{d}x \\
&= x\,(\mathrm{e}^x + 1) - (\mathrm{e}^x + x) + C \\
&= x\,\mathrm{e}^x - \mathrm{e}^x + C
\end{aligned}$$

In (10.4.1), by putting $u = f(x)$ and $v = g(x)$ so that $\mathrm{d}u = f'(x)\,\mathrm{d}x$ and $\mathrm{d}v = g'(x)\,\mathrm{d}x$, we get

$$\int u\,\mathrm{d}v = uv - \int v\,\mathrm{d}u \tag{10.4.2}$$

In applying (10.4.2), we have to choose suitable u and $\mathrm{d}v$. From the chosen $\mathrm{d}v$, we have to find v. This is done by integration. For example, suppose that $\mathrm{d}v = \mathrm{e}^x\,\mathrm{d}x$, which means that

$$\frac{\mathrm{d}v}{\mathrm{d}x} = \mathrm{e}^x.$$

Integrating, we get

$$v = \mathrm{e}^x + C.$$

To apply the formula, we only need to take a suitable v (see the solution and the remark of the preceding example). Below we redo the example using *integration by part*, that is, using (10.4.2).

Example Find $\int x\,\mathrm{e}^x\,\mathrm{d}x$.

Solution Put $u = x$ and $\mathrm{d}v = \mathrm{e}^x\,\mathrm{d}x$. Then we have $\mathrm{d}u = \mathrm{d}x$ and we can take $v = \mathrm{e}^x$. From these we get

$$\begin{aligned}
\int x\,\mathrm{e}^x\,\mathrm{d}x &= \int u\,\mathrm{d}v && \text{Substitution} \\
&= uv - \int v\,du && \text{Integration by parts} \\
&= x\,\mathrm{e}^x - \int \mathrm{e}^x\,\mathrm{d}x && \text{Back substitution} \\
&= x\,\mathrm{e}^x - \mathrm{e}^x + C
\end{aligned}$$

□

A Guide for Integration by Parts Treat the integrand as a product of two functions. Choose u to be one of the two functions such that

- the other function can be integrated easily—choose $\mathrm{d}v$ = (the other function) $\mathrm{d}x$;
- the new integral $\int v\,\mathrm{d}u$ is easier to find than the original integral $\int u\,dv$.

Example Find $\int x\,\cos x\,\mathrm{d}x$.

Explanation The integrand is a product of two functions. There are two options for u and $\mathrm{d}v$.

- Put $u = x$ and $dv = \cos x\,dx$. Then we have $du = dx$ and we can take $v = \sin x$. Note that $\int v\,du = \int \sin x\,dx$ is easy to find (the method works).
- Put $u = \cos x$ and $dv = x\,dx$. Then we have $du = -\sin x\,dx$ and we can take $v = \frac{1}{2}x^2$. Note that $\int v\,du = \int -\frac{1}{2}x^2 \sin x\,dx$ which is even more complicated than the original integral.

Solution Put $u = x$ and $dv = \cos x\,dx$. Then we have $du = dx$ and we can take $v = \sin x$. From these we get

$$\begin{aligned}\int x\cos x\,dx &= x\sin x - \int \sin x\,dx \qquad \text{Integration by parts}\\ &= x\sin x + \cos x + C\end{aligned}$$

□

Example Find $\int \ln x\,dx$.

Explanation The integrand can be written as $\ln x \cdot 1$, a product of two functions. To choose u and dv, there is only one plausible way, namely $u = \ln x$ and $dv = dx$. Readers may try to see what happens if we choose $u = 1$ and $dv = \ln x\,dx$.

Solution Put $u = \ln x$ and $dv = dx$. Then we have $du = \frac{1}{x}\,dx$ and we can take $v = x$. From these we get

$$\begin{aligned}\int \ln x\,dx &= \ln x \cdot x - \int x \cdot \frac{1}{x}\,dx \qquad \text{Integration by parts}\\ &= x\ln x - \int 1\,dx\\ &= x\ln x - x + C\end{aligned}$$

□

10.5 More Applications of Definite Integrals

In economics, we have the concepts of *consumers' surplus* and *producers' surplus*. These two concepts are defined in terms of definite integrals.

Consumers' and Producers' Surplus

Let $p = D(q)$ and $p = S(q)$ be respectively the demand and supply equations for a certain product. The quantity q_0 at which $D(q) = S(q)$ is called the *equilibrium quantity* and the corresponding price p_0 is called the *equilibrium price*.

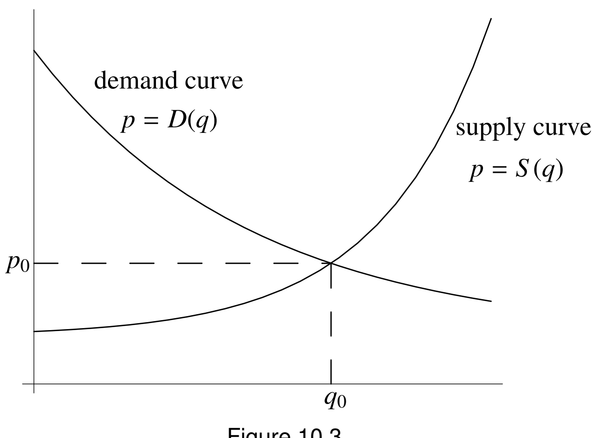
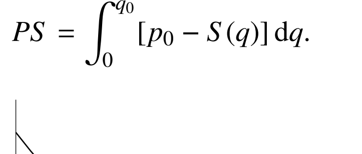


Figure 10.3

Note that (q_0, p_0) is the intersection point of the demand curve and the supply curve.

The consumers' surplus (denoted by CS) and producers' surplus (denoted by PS) under market equilibrium are defined as follows:

$$CS = \int_0^{q_0} [D(q) - p_0]\, dq, \qquad PS = \int_0^{q_0} [p_0 - S(q)]\, dq.$$



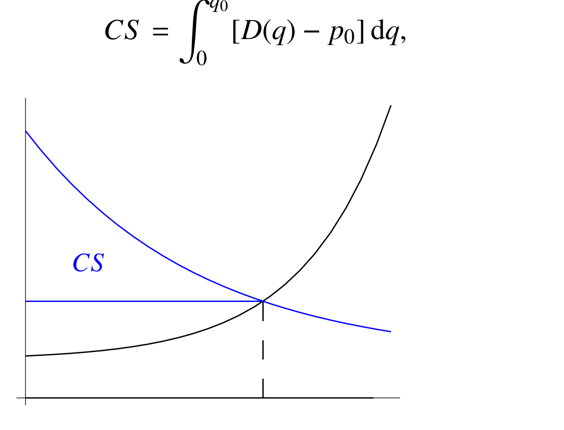


Figure 10.4(a)

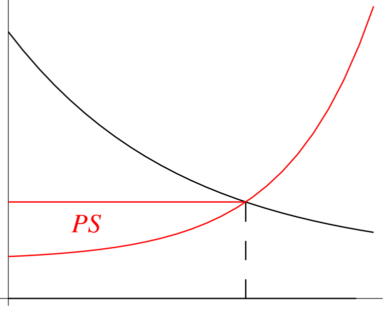


Figure 10.4(b)

Example Find the consumers' surplus and producers' surplus if the demand and supply equations are

$$p = D(q) = 20 - \frac{q}{20}, \qquad p = S(q) = 2 + \frac{q^2}{5000}.$$

Solution First we find the intersection point (q_0, p_0) of the demand curve $p = D(q)$ and the supply curve $p = S(q)$. Solving $D(q) = S(q)$ (noting that $q > 0$)

$$\begin{aligned} 20 - \frac{q}{20} &= 2 + \frac{q^2}{5000} && (q > 0) \\ \frac{q^2}{5000} + \frac{q}{20} - 18 &= 0 && (q > 0) \\ q^2 + 250q - 90000 &= 0 && (q > 0) \\ (q - 200)(q + 450) &= 0 && (q > 0) \end{aligned}$$

we get $q_0 = 200$ and so $p_0 = D(200) = 10$.

The consumers' surplus is

$$\begin{aligned} CS &= \int_0^{200} [D(q) - 10]\, dq \\ &= \int_0^{200} \left(20 - \frac{q}{20} - 10\right) dq \\ &= \int_0^{200} \left(10 - \frac{q}{20}\right) dq \\ &= \left[10q - \frac{q^2}{40}\right]_0^{200} \\ &= 1000. \end{aligned}$$

The producers' surplus is

$$\begin{aligned} PS &= \int_0^{200} [10 - S(q)]\,\mathrm{d}q \\ &= \int_0^{200} \left[10 - \left(2 + \frac{q^2}{5000}\right)\right] \mathrm{d}q \\ &= \int_0^{200} \left(8 - \frac{q^2}{5000}\right) \mathrm{d}q \\ &= \left[8q - \frac{q^3}{15000}\right]_0^{200} \\ &= \frac{3200}{3}. \end{aligned}$$

□

Probability

To consider probabilities, the simplest method is to count, that is, to do addition. This works for the case where the sample space is finite. However, if the sample space is infinite, we can't count. To define probabilities, we use definite integration which can be considered as a generalization of addition. Below we give a very brief introduction to probabilities of events for continuous random variables.

Definition A variable whose values depend on the outcome of a random process is called a *random variable*.

Example

(1) Suppose a die is rolled and X_1 is the number that turns up. Then X_1 is a random variable with values in $\{1, 2, 3, 4, 5, 6\}$.

(2) The life (in months) of a certain computer part is a random variable X_2 with values in $[0, \infty)$.

Note that the values that X_1 can take are discrete whereas X_2 can take any value in the interval $[0, \infty)$. For this we say that X_1 is a *discrete* random variable and X_2 a *continuous* random variable.

Definition Let X be a discrete random variable with values in $\{x_1, x_2, \ldots, x_n\}$. A *probability function* of X is a function f with domain $\{x_1, x_2, \ldots, x_n\}$ such that

(1) $0 \le f(x_i)$ for all $i = 1, \ldots, n$;

(2) $f(x_1) + \cdots + f(x_n) = 1$.

Example Suppose a die is rolled and X is the number that turns up. If the die is fair, then the probability of getting any one of the six numbers is $\frac{1}{6}$. Thus we have the following probability function of X:

$$f(i) = \frac{1}{6} \quad \text{for } 1 \le i \le 6.$$

More generally, if the die is not fair, then the probability function g of X is given by

$$g(i) = w_i \quad \text{for } 1 \le i \le 6,$$

where $0 < w_i < 1$ and $w_1 + \cdots + w_6 = 1$.

From the probability function g, we can find (for example) the probability of getting an odd number:

$$P(X \text{ is odd}) = w_1 + w_3 + w_5 = \sum_{i \text{ is odd}} g(i),$$

where (X is odd) denotes the event that the number X that turns up is odd and $P(X$ is odd) denotes the probability of the event (X is odd).

Probabilities of Events for Discrete Random Variables Suppose that X is a discrete random variable with values in the set $\{x_1, \dots, x_n\}$ and that f is a probability function of X.

- An *event* for X is a subset of $\{x_1, \dots, x_n\}$.
- The probability of an event E, denoted by $P(E)$, is the number given by

$$P(E) = \sum_{x_i \in E} f(x_i).$$

For continuous random variable, instead of taking sums, we consider integration.

Definition Let X be a continuous random variable with values in $[a, \infty)$. A *probability function* of X is a function f with domain $[a, \infty)$ such that

(1) $0 \le f(x)$ for all $x \in [a, \infty)$;

(2) $\displaystyle\int_a^\infty f(x)\,\mathrm{d}x = 1.$

Remark $\displaystyle\int_a^\infty f(x)\,\mathrm{d}x$ is called an *improper integral* and is defined by

$$\int_a^\infty f(x)\,\mathrm{d}x = \lim_{R\to\infty} \int_a^R f(x)\,\mathrm{d}x$$

provided that the limit exists. For example,

$$\begin{aligned}\int_1^\infty \frac{1}{x^2}\,\mathrm{d}x &= \lim_{R\to\infty} \int_1^R x^{-2}\,\mathrm{d}x \\ &= \lim_{R\to\infty} \left[\frac{x^{-1}}{-1}\right]_1^R \\ &= \lim_{R\to\infty} \left(-R^{-1} - (-1)\right) \\ &= 1.\end{aligned}$$

Probabilities of Events for Continuous Random Variables Suppose that X is a continuous random variable with values in the interval $[a, \infty)$ and that f is a probability function of X.

- An *event* for X is a "nice" subset of $[a, \infty)$, where "nice" means that the integral of f over that subset "can be found". In most cases, we consider events that are intervals contained in $[a, \infty)$; such events are represented by $(\alpha \le X \le \beta)$, where $a \le \alpha < \beta \le \infty$.

- The probability of an event $(\alpha < X < \beta)$, denoted by $P(\alpha < X < \beta)$, is the number given by

$$P(\alpha < X < \beta) = \int_{\alpha}^{\beta} f(x)\,dx.$$

Remark It doesn't matter whether we include the endpoints α and β. For a continuous random variable X, the probability that X equals a specific value is 0.

Example The life (in months) of a certain computer part has probability function given by

$$f(x) = \frac{1}{18}e^{-\frac{x}{18}}, \quad x \in [0, \infty).$$

Find the probability that a randomly selected component will last

(1) between 1 year and $1\frac{1}{2}$ years;

(2) at most 6 months;

(3) more than 2 years.

Explanation In this example, the random variable X is the life of a computer part. Note that X has values in $[0, \infty)$. The given function f is a probability function for X. This can be checked as follows:

$$\begin{aligned}
\int_0^{\infty} f(x)\,\mathrm{d}x &= \int_0^{\infty} \frac{1}{18}e^{-\frac{x}{18}}\,\mathrm{d}x \\
&= \lim_{R\to\infty} \int_0^{R} \frac{1}{18}e^{-\frac{x}{18}}\,\mathrm{d}x \\
&= \lim_{R\to\infty} \left[\frac{1}{18} \cdot \frac{1}{\frac{-1}{18}}e^{-\frac{x}{18}}\right]_0^R \\
&= \lim_{R\to\infty} \left(-e^{-\frac{R}{18}} + 1\right) \\
&= 1.
\end{aligned}$$

Solution

(1) The given event is $(12 < X < 18)$. The probability of the event is

$$\begin{aligned}
P(12 < X < 18) &= \int_{12}^{18} \frac{1}{18}e^{-\frac{x}{18}}\,\mathrm{d}x \\
&= \left[-e^{-\frac{x}{18}}\right]_{12}^{18} \\
&= -e^{-1} + e^{-\frac{2}{3}} \\
&\approx 0.146.
\end{aligned}$$

(2) The given event is $(X \le 6)$. The probability of the event is the same as that of $(0 < X < 6)$.

$$\begin{aligned}
P(0 < X < 6) &= \int_{0}^{6} \frac{1}{18}e^{-\frac{x}{18}}\,\mathrm{d}x \\
&= \left[-e^{-\frac{x}{18}}\right]_{0}^{6} \\
&= 1 - e^{-\frac{1}{3}} \\
&\approx 0.283.
\end{aligned}$$

(3) The given event is $(X > 24)$, that is, $(24 < X < \infty)$. The probability of the event is

$$\begin{aligned}
P(24 < X < \infty) &= \int_{24}^{\infty} \frac{1}{18} e^{-\frac{x}{18}} \, dx \\
&= \lim_{R\to\infty} \int_{24}^{R} \frac{1}{18} e^{-\frac{x}{18}} \, dx \\
&= \lim_{R\to\infty} \left[-e^{-\frac{x}{18}} \right]_{24}^{R} \\
&= \lim_{R\to\infty} \left(e^{-\frac{4}{3}} - e^{-\frac{R}{18}} \right) \\
&= e^{-\frac{4}{3}} \\
&\approx 0.264.
\end{aligned}$$

□

Appendix A

Answers

Exercise 0.1

1. (a) x^3 (b) $\frac{y^2z^3}{x}$ (c) $\frac{y^{12}}{x^8}$ (d) $\frac{y^2}{8x^4}$

Exercise 0.2

1. (a) $4x^2 + 12x + 9$ (b) $9x^2 - 6xy + y^2$ (c) $x^2 - 9y^2$
   (d) $x^2 + 7xy + 12y^2$ (e) $4x - 12\sqrt{x} + 9$ (f) $x - 25$
2. (a) $(x-3)(x-4)$ (b) $(x+3)(x-2)$ (c) $(x+4)^2$
   (d) $(3x+1)(3x+2)$ (e) $(3x-1)^2$ (f) $5(x+1)(x-1)$
   (g) $3(x-3)^2$ (h) $2(x-2)(x-4)$
3. (a) $\frac{(x+2)}{(x-4)}$ (b) $\frac{-(x+4)}{(x+2)}$
   (c) $\frac{1}{2(x+1)^2}$ (d) $\frac{-1}{x(x+h)}$

Exercise 0.3

1. (a) $\frac{6}{5}$ (b) 2 (c) $\frac{b^2}{a-b}$ (d) $\frac{abc}{b-a}$

Exercise 0.4

1. (a) 0, 1 (b) 1, $\frac{-2}{3}$ (c) 3, $\frac{5}{4}$
   (d) $-\sqrt{2}$ (e) no solution (f) 0, $\frac{7 \pm \sqrt{37}}{2}$
2. −2, 6
3. 14

Exercise 0.5

1. (a) $(x-1)(x-3)(x+4)$ (b) $(x-1)(x-3)(2x+1)$
   (c) $(x-1)^2(2x+3)$ (d) $(x-1)(x^2-4x+7)$
2. (a) $-\frac{3}{2}$, 1, 5 (b) 1, $\frac{-1 \pm \sqrt{5}}{2}$ (c) 3

Exercise 0.6

1. (a) $x \le \frac{17}{9}$ (b) $x \ge \frac{6-\sqrt{3}}{2-\sqrt{3}}$
   (c) $x > 1$ (d) $x < -\frac{3}{2}$

Exercise 0.7

1. (a) $3x + 2y = 0$ (b) $2x - y - 11 = 0$ (c) $3x + y - 7 = 0$
   (d) $2x - y + 8 = 0$ (e) $3x - y + 1 = 0$ (f) $y + 1 = 0$

Exercise 0.8

1. (a) 5 (b) $\sqrt{65}$ (c) 13 (d) $5\sqrt{5}$
2. (a) $r = \sqrt{3}$ $C = (0, 2)$ (b) $r = 3$ $C = (-2, 1)$ (c) $r = \frac{\sqrt{3}}{2}$ $C = (-1, \frac{1}{2})$
3. (a) 2 (b) $\frac{1}{\sqrt{2}}$ (c) $\frac{2}{\sqrt{5}}$

Exercise 0.9

1. (a) x-intercepts: $(-6, 0), (2, 0)$, y-intercept: $(0, -12)$, vertex: $(-2, -16)$
   (b) x-intercepts: $(3 \pm \sqrt{2}, 0)$, y-intercept: $(0, -7)$, vertex: $(3, 2)$
   (c) no x-intercept, y-intercept: $(0, 7)$, vertex: $(-\frac{1}{2}, \frac{13}{2})$

Exercise 0.10

1. $8\,cm \times 6\,cm$

Exercise 1.1

1. (a) $\{2, 3, 5, 7\}$
   (b) $\{2, 4, 6, 8, 10\}$
   (c) $\{2\}$
   (d) $\{1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 13, 17, 19\}$
   (e) $\{1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 12, 14, 16, 18\}$
   (f) $\{2, 3, 4, 5, 6, 7, 8, 10, 11, 12, 13, 14, 16, 17, 18, 19\}$
   (g) $\{1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 16, 17, 18, 19\}$
   (h) $\{2\}$
   (i) $\{2, 4, 6, 8, 10\}$
   (j) $\{2, 3, 4, 5, 6, 7, 8, 10, 12, 14, 16, 18\}$
   (k) $\{1, 4, 6, 8, 9, 10\}$
   (l) $\{12, 14, 15, 16, 18\}$
2. (a) False (b) False (c) True

Exercise 1.2

1. (a) $\{\sqrt{2}, -\sqrt{2}\}$ (b) $\{\sqrt{2}\}$ (c) $\emptyset$
2. (a) $[3, 5]$ (b) $[1, 9)$ (c) $(1, 5)$
   (d) $\{5\}$ (e) $\{1\}$ (f) $[3, 5) \cup (5, 9)$
   (g) $\{5\}$ (h) $[1, \infty)$ (i) $\{5\}$

Exercise 1.3

1. (a) $x \leq -\frac{7}{5}$ (b) $x < -\frac{29}{3}$
   (c) $x > 2$ (d) $x \leq -\frac{7}{2}$ or $x \geq \frac{5}{11}$
   (e) $-1 < x < 3$ (f) $x < \frac{3 - \sqrt{41}}{4}$ or $x > \frac{3 + \sqrt{41}}{4}$
   (g) no solution (h) $x \leq -\frac{3}{2}$ or $x > 4$
   (i) $-7 < x < 4$

2. (a) $x(x+5)(2x-3)$ (b) $(x-1)(x+1)(2x+3)$
   (c) $(x-2)(x^2+x+1)$ (d) $x(x-5)(x-1)(x+3)$
   (e) $(x-2)(x+1)(x-1)^2$ (f) $(x-1)^2(x^2+x+2)$
3. (a) $-\frac{3}{2} < x < \frac{9}{5}$ or $x > 4$ (b) $x \le 3$
   (c) $x < -2$ or $1 < x < 3$ (d) $-3 \le x \le \frac{3}{2}$ or $x \ge 2$
   (e) $x > 3$ (f) $x \le -1$
   (g) $x < -4$ or $-2 < x < 1$ or $x > 3$ (h) $-\frac{5}{3} \le x \le \frac{3}{2}$

Exercise 2.1

1. (a) $-\frac{3}{8}$ (b) $-\frac{6}{65}$ (c) $\frac{a-4}{a^2+2a+5}$
   (d) $\frac{\sqrt{a}-5}{a+4}$ (e) $\frac{a^2-5}{a^4+4}$ (f) $\frac{a-5}{a^2+4} - \frac{4}{5}$
2. (a) $\frac{1}{2}$ (b) $\frac{2}{3}$ (c) $\frac{3}{4\sqrt{2}}$
   (d) $1 - \frac{1}{a} + \sqrt{a}$ (e) $\frac{(a^2+1)|a|}{a^2+2}$
3. (a) $a^2 + 2ab + b^2 - 3a - 3b + 4$ (b) $-1 + h$
   (c) $2a - 3 + h$

Exercise 2.2

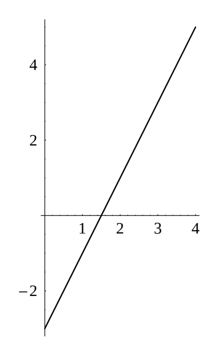
1. (a) $\mathbb{R}$ (b) $\mathbb{R} \setminus \{\frac{-6}{5}\}$ (c) $\mathbb{R} \setminus \{-\sqrt{5}, \sqrt{5}\}$
   (d) $\mathbb{R} \setminus \{-1, 3\}$ (e) $(\frac{3}{2}, \infty)$ (f) $[-3, \infty) \setminus \{\frac{1}{2}\}$
   (g) $[-\frac{5}{2}, \infty) \setminus \{-1, 1\}$ (h) $(-\infty, -5) \cup (2, \infty)$
2. (a) $[-5, \infty)$ (b) $[-4, \infty)$ (c) $\mathbb{R} \setminus \{0\}$
   (d) $\mathbb{R} \setminus \{3\}$ (e) $(0, \infty)$ (f) $(-\infty, -\frac{1}{5}] \cup (0, \infty)$
   (g) $(-\infty, -\frac{1}{4}] \cup (0, \infty)$
3. $A(w) = 14w - w^2$, dom $= (0, 14)$, range $= (0, 49]$

Exercise 2.3

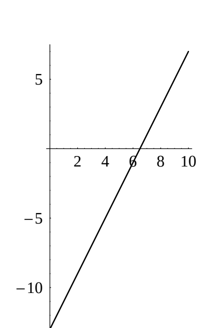
1. x-intercept: $(\sqrt{2}, 0)$, $(-\sqrt{2}, 0)$, y-intercept: $(0, \frac{2}{\sqrt{3}})$, $(0, -\frac{2}{\sqrt{3}})$
2. $a = 1, b = 1, c = -6$
3. (a) y-intercept: $(0, 5)$, no x-intercept
4. (a) $(2, -1)$, $(\frac{2}{5}, \frac{11}{5})$ (b) $(2, -1)$, $(\frac{2}{3}, \frac{5}{3})$
   (c) $\{(2, 1), (2, -1), (-2, 1), (-2, -1)\}$
5. $a = \pm\sqrt{3}$

Exercise 2.4

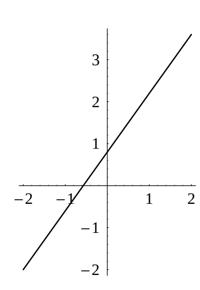
1. (a)



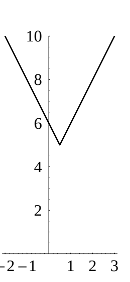
(b)



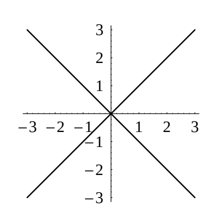
(c)



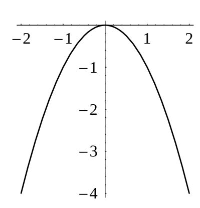
(d)



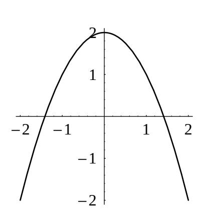
(e)



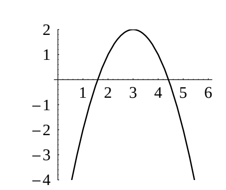
(f)



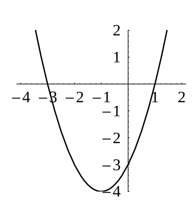
(g)



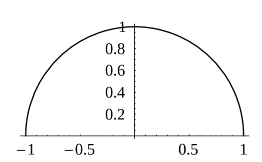
(h)



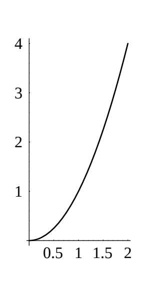
(i)



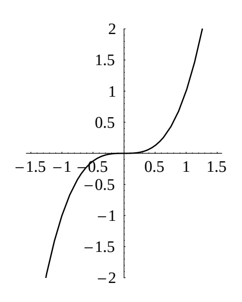
(j)



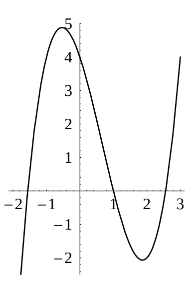
(k)



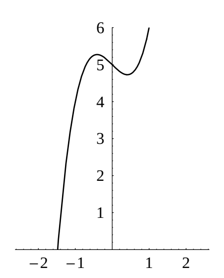
2. (a)



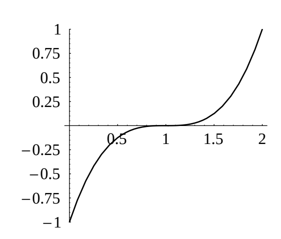
(b)



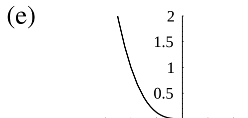
(c)

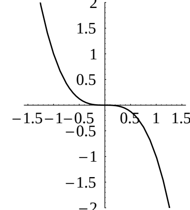


(d)



(e)

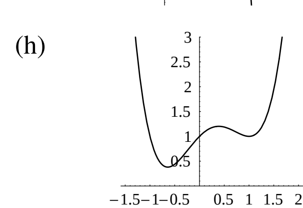




(f)

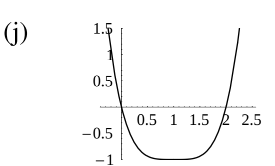
(g)

(h)

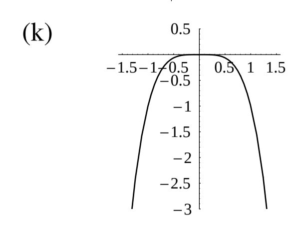


(i)

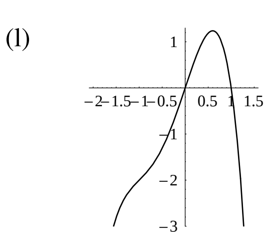
(j)



(k)



(l)



3. $\left(\frac{1-\sqrt{13}}{2}, \frac{-1-\sqrt{13}}{6}\right)$, $\quad \left(\frac{1+\sqrt{13}}{2}, \frac{-1+\sqrt{13}}{6}\right)$

4. (a) 1 second

   (b) $\frac{9}{5}$ meters

5. (a) $R = 1600000 + 20000n - 500n^2$, $\quad$ domain $= \{0, 1, 2, 3, \ldots, 80\}$

   (b) $n = 20$, $\quad$ \$1800000

Exercise 2.5

1. (a) 5 $\quad$ (b) 3 $\quad$ (c) $x^2 + 2x + 2$

   (d) $x^2 + 2$ $\quad$ (e) $a^4 + 2a^2 + 2$ $\quad$ (f) $a + 2$

2. (a) $f(x) = x^2 + 1,\ g(x) = x^{\frac{1}{2}}$ $\quad$ (b) $f(x) = x + 1,\ g(x) = x^{-1}$

Exercise 2.6

1. (a) Yes $\quad$ (b) No

2. (a) $f^{-1}(x) = \dfrac{x+2}{3}$ $\quad$ (b) $f^{-1}(x) = (x-3)^{\frac{1}{5}}$

   (c) $f^{-1}(x) = \dfrac{(x-1)^7}{128}$ $\quad$ (d) $f^{-1}(x) = \sqrt[3]{\dfrac{x^3+1}{2}}$

Exercise 2.7

1. (a) $\{\frac{5 \pm \sqrt{33}}{2}\}$ $\quad$ (b) $\{-1, 2\}$ $\quad$ (c) $\{0\}$

   (d) $\mathbb{R} \setminus \{2, -2\}$ $\quad$ (e) $\{2\}$ $\quad$ (f) $\{5\}$

   (g) $\emptyset$ $\quad$ (h) $\{4\}$ $\quad$ (i) $\{1, 2\}$

2. (a) $P(q) = 8q - q^2 - 12$
   (b) $q = 2, 6$
3. (a) 106 ft
   (b) 51 mph
4. The sides are $3, 4, 5$

Exercise 3.1

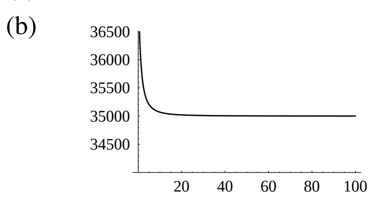
1. Same result, provided that the lengths tend to zero.
2. (a) Same result even for arbitrary point.
   (b) Same result, provided that the lengths tend to zero.

Exercise 3.2

1. (a) 0 (b) 7
   (c) $\frac{3}{2}$ (d) ∞, does not exist
   (e) 0 (f) Does not exist.
2. (a) (i) \$51007.53 (ii) \$51009.22
   (b) $\$50000(1 + \frac{1}{50n})^n$
   (c) Limit exists, approximately \$51010.07 (exact value is $50000e^{\frac{1}{50}}$).
3. (a) Limit exists, approximately 2.718 (exact value is e).
   (b) Limit exists, approximately 7.389 (exact value is e^2).
   (c) The limit is 1.
4. (a) Limit exists, the value is nonnegative.

Exercise 3.3

1. (a) 0 (b) 15
   (c) 0 (d) ∞, does not exist
   (e) 0 (f) 1
   (g) ∞, does not exist (h) 1
   (i) -1 (j) Does not exist
2. (a) 0 (b) The concentration will drop to 0 in the long run.
3. (a) 35000
   (b)



The population decreases from initial population 37500 to 35000.

4. (a) 0 (b) $\frac{1}{2}$ (c) 0

Exercise 3.4

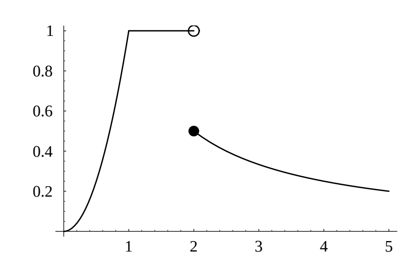
1. (a) 0 (b) ∞, does not exist.
   (c) $-\infty$, does not exist. (d) Does not exist.
   (e) $-\infty$, does not exist. (f) 2

Exercise 3.5

1. (a) 6 (b) 216 (c) $\frac{25}{49}$
   (d) $\frac{1}{\sqrt{2}}$ (e) $\frac{1}{8}$ (f) $\frac{4}{3}$
   (g) 0 (h) $\frac{6}{7}$ (i) ∞, does not exist.
   (j) 0 (k) $\frac{1}{4}$ (l) -3
2. (a) 4 (b) $3x^2$
   (c) $-\frac{1}{x^2}$ (d) $\frac{1}{2\sqrt{x}}$

Exercise 3.6

1. (a)



   (b) 2
2. (a) $[0, \infty) \setminus \{1\}$
   (b) -6
   (c) Yes, define $f(1) = -6$.
3. (a) $\mathbb{R} - \{0\}$
   (b) Does not exist.
   (c) No
4. (a) $-1 < x < 2$ or $x > 2$
5. (a) $p(1)$ and $p(2)$ have opposite signs
   (b) Closer to one, the solution lies between 1 and 1.5

Exercise 4.1

1. (a) $4x$ (b) $3x^2 - 3$
   (c) $4x^3$ (d) $-\frac{2}{x^3}$

Exercise 4.2

1. (a) 0 (b) $18x^8 + 3$ (c) $2x + 5$
   (d) $2x - 1$ (e) $28 - 24x$ (f) $6x(x^2 + 5)^2$
   (g) $-92x^{-5}$ (h) x^{-2} (i) $\frac{2}{(x+1)^2}$
   (j) $1 + \frac{1}{2}x^{-\frac{1}{2}}$
2. (a) -1 (b) $-\frac{11}{8}$ (c) -1
   (d) $8\pi - \frac{1}{2}$ (e) 18 (f) -6
   (g) -10
3. (a) 72 (b) $y = -1$ (c) $(-1, -1),\ (0, 2),\ (1, -1)$

4. (a) $x\cos x + \sin x$ (b) $\dfrac{\cos x}{x+1} - \dfrac{\sin x}{(x+1)^2}$

   (c) $\dfrac{2x}{\sin x} - \dfrac{(x^2+1)\cos x}{\sin^2 x}$ (d) $x(x+2)\cos x + 2(x+1)\sin x$

5. (c) $nf(x)^{n-1}\dfrac{\mathrm{d}}{\mathrm{d}x}f(x)$

6. (a) $2(3x^2+10x)(x^3+5x^2-2)$ (b) $6x(x^2+5)^2$

Exercise 4.3

1. (a) $6x-6$ (b) -4 (c) $\dfrac{3}{4}x^{-\frac{1}{2}} - \dfrac{1}{4}x^{-\frac{3}{2}}$

   (d) $6x^{-4} - 4x^{-3}$ (e) $6x(5x^3+2)$

2. (a) 50 (b) -22 (c) 18

3. (a) a_0, a_1 (b) $a_n n(n-1)\cdots 1$, 0

Exercise 5.1

1. (a) $[\frac{5}{4}, \infty)$ (b) $[-1, 1]$

   (c) $(-\infty, -7]$, $[3, \infty)$ (d) $(-\infty, \frac{-3-\sqrt{13}}{2}]$, $[\frac{-3+\sqrt{13}}{2}, \infty)$

   (e) $[-2, -1]$, $[2, \infty)$ (f) $(-\infty, -2]$, $[2, \infty)$

2. (a) $\dfrac{7}{2}$ (local maximizer)

   (b) 0 (neither), $\dfrac{3}{2}$ (local minimizer)

   (c) -3 (local maximizer), 0 (neither), 3 (local minimizer)

   (d) -2 (local maximizer), 0 (local minimizer)

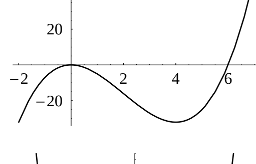
3. (a) none (b) $(2, \infty)$

   (c) $(0, \frac{4}{5})$, $(1, \infty)$ (d) $(0, \infty)$

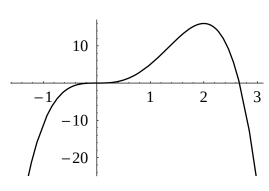
4. (a) $-\dfrac{3}{2}$ (b) 3



5. (a)

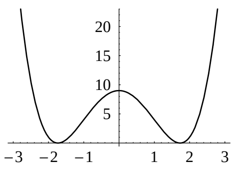


(b)

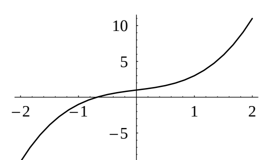




(c)



(d)



Exercise 5.2

1. (a) minimum: -10, maximum: 82

   (b) minimum: 0, maximum: 58

   (c) minimum: -4, maximum: $\dfrac{17}{16}$

2. 25, 25

3. $x = 6$, $y = 3$

4. 10 units by 10 units

5. 100 m by 150 m
6. $(2 + 2\sqrt{6})$ in. by $(3 + 3\sqrt{6})$ in.
7. $\frac{3}{4}\,s$, maximum height = 14 ft.
8. $\frac{1}{4}$ month, maximum height = $\frac{1}{4}$ meter
9. Maximum profit: $30000, produce 1500 pieces, price for each piece: $200.

Exercise 6.1

1. (a) 0 (b) -6
2. $\frac{1}{4}$

Exercise 6.2

1. (a) $\frac{x^2}{2}$, yes (b) x, yes (c) $\frac{x^6}{6}$, yes
   (d) $x^2 + x$, yes (e) $\sqrt{x}$, contained in $(0, \infty)$ (f) $\frac{2}{3}x^{\frac{3}{2}}$, contained in $[0, \infty)$
2. $\frac{x^5}{5}$ (a) $\frac{1}{5}$ (b) $\frac{242}{5}$ (c) $\frac{243}{5}$ (d) $\frac{1701}{5}$

Exercise 6.3

1. (a) $\frac{x^6}{3} + C$ (b) $3x - 8\sqrt{x} + C$
   (c) $\frac{1}{8}x^8 - \frac{3}{2}x^2 + 2x + C$ (d) $\frac{1}{3}x^3 - \frac{2}{3}x^{\frac{3}{2}} + 3x + C$
   (e) $-\frac{4}{3\sqrt{x}} + C$ (f) $-\frac{3}{4}x^4 + \frac{17}{3}x^3 - \frac{13}{2}x^2 + 2x + C$
   (g) $\frac{1}{5}x^5 - 2x^3 + 9x + C$ (h) $x - \frac{1}{x} + C$
2. (a) $\frac{81}{2}$ (b) 0 (c) -30 (d) $\frac{42}{5}$
   (e) $\frac{84}{5}$ (f) 22 (g) $\frac{18\sqrt{2} - 12}{5}$ (h) 12

Exercise 6.4

1. (a) $\frac{81}{4}$ (b) $\frac{11}{3}$ (c) $\frac{37}{2}$ (d) $\frac{16}{3}$
2. (a) $\frac{1}{6}$ (b) $\frac{125}{3}$ (c) $\frac{37}{12}$ (d) 64
3. $\frac{1}{3}x^3 + x + \frac{2}{3}$
4. $\frac{1}{12}x^4 - \frac{1}{6}x^3 - x^2 + \frac{1}{12}x + 1$
5. $\frac{3375}{2}$ liters

Exercise 7.1

1. (a) $\frac{3\pi}{2}$ (b) $\frac{7\pi}{6}$ (c) $\frac{7\pi}{4}$ (d) $\frac{25\pi}{6}$
2. (a) 30° (b) 135° (c) 450° (d) 1260°

Exercise 7.2

1. (a) $\frac{\sqrt{3}}{2}$ (b) $-\frac{1}{2}$ (c) $-\sqrt{3}$

   (d) $-\frac{1}{\sqrt{2}}$ (e) $-\frac{1}{\sqrt{2}}$ (f) 1

2. (a) 1 (b) 2

Exercise 7.3

1. (a) $-5\sin x - 2$ (b) $-2\sec^2 x$

   (c) $\cos x - 2x$ (d) $x^2\cos x + 2x\sin x$

   (e) $-2\sin x\cos x$ (f) $\sin x\sec^2 x$

   (g) $\cos^2 x - \sin^2 x$ (h) $\frac{-(x^3+1)\sin x - 3x^2\cos x}{(x^3+1)^2}$

   (i) $2(x+\cos x)(1-\sin x)$ (j) $2(\cos^2 x - \sin^2 x)$

2. (a) $2\cos x\sin x,\quad 3\cos x\sin^2 x,\quad 4\cos x\sin^3 x$

   (b) $n\cos x\sin^{n-1} x$

3. (a) $2\cos 2x,\quad -2\sin 2x$

   (b) $3\cos 3x,\quad -3\sin 3x$

   (c) $n\cos nx,\quad -n\sin nx$

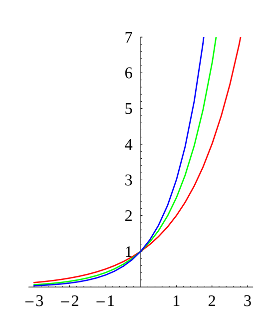
4. (a) $a\cos(ax+b),\quad -a\sin(ax+b)$

   (b) $-a^2\sin(ax+b),\quad -a^2\cos(ax+b)$

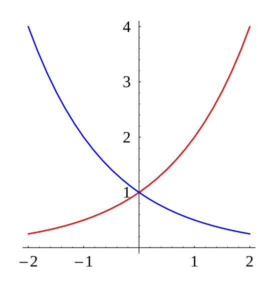
   (c)
$$f^{(n)}(x) = \begin{cases} (-1)^{\frac{j-1}{2}} a^n\cos(ax+b) & \text{if } n = 4i+j,\ j = 1,3, \\ (-1)^{\frac{j}{2}} a^n\sin(ax+b) & \text{if } n = 4i+j,\ j = 2,4. \end{cases}$$
$$g^{(n)}(x) = \begin{cases} (-1)^{\frac{j+1}{2}} a^n\sin(ax+b) & \text{if } n = 4i+j,\ j = 1,3, \\ (-1)^{\frac{j}{2}} a^n\cos(ax+b) & \text{if } n = 4i+j,\ j = 2,4. \end{cases}$$

Exercise 8.1

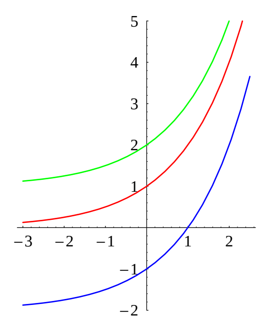
1. (a)



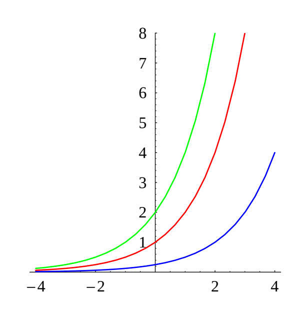
(b)



(c)



(d)

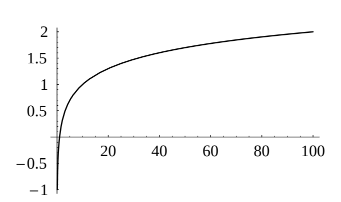


red= 1st, green=2nd, blue=3rd

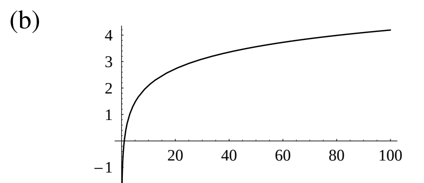
2. (a) domain = $\mathbb{R}$, range = $(1, \infty)$
   (b) domain = $\mathbb{R}$, range = $(0, 1)$
   (c) domain = $\mathbb{R} \setminus \{0\}$, range = $(-\infty, -1) \cup (0, \infty)$
3. (a) $-1, 3$ (b) -2 (c) no solution (d) $0, 2$

Exercise 8.2

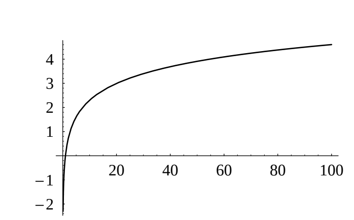
1. (a) $\log_9 81 = 2$ (b) $\log_4 2 = \frac{1}{2}$ (c) $\log_2 \frac{1}{2} = -1$
2. (a) $2^3 = 8$ (b) $9^{\frac{3}{2}} = 27$ (c) $e^0 = 1$
3. (a) $\frac{2}{9}$ (b) 1
4. (a)



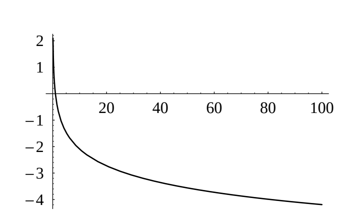
(b)



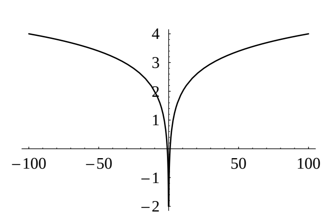
(c)



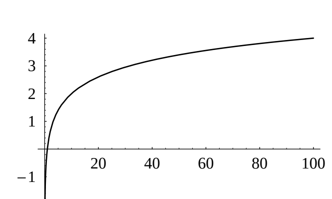
(d)



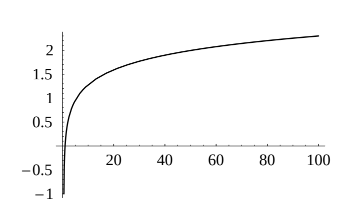
(e)



(f)



(g)



5. (a) domain = $\mathbb{R}$, range = $(1, \infty)$
   (b) domain = $\mathbb{R}$, range = $[1, \infty)$
   (c) domain = $\mathbb{R} - \{0\}$, range = $\mathbb{R}$
   (d) domain = $(0, \infty)$, range = $\mathbb{R}$
   (e) domain = $(\frac{1}{2}, \infty)$, range = $\mathbb{R}$
   (f) domain = $(-\infty, -2) \cup (2, \infty)$, range = $\mathbb{R}$
6. (a) $\{4\}$ (b) $\{32\}$ (c) $\{3\}$
   (d) $\{1, 2\}$ (e) $\{3\}$ (f) $\{7\}$
7. (a) 2.09 (b) 8.64 (c) 2.69×10^{-14}
   (d) 1.02 (e) 4.71 (f) 1.99
8. (a) 30.81 years (b) 30.55 years (c) 30.50 years (d) 30.47 years

Exercise 8.3

1. (a) $6x^2 - 4e^x$ (b) $\frac{1}{x}$
   (c) $e^x + \frac{1}{x}$ (d) $2x + \frac{2}{x}$
   (e) $e^x + \frac{1}{2\sqrt{x}}$ (f) $\frac{1}{2x}$
   (g) $e^x(\sin x + \cos x)$ (h) $\frac{\cos x}{x} - \sin x \ln x$
   (i) $(x+1)^2 e^x$ (j) $x + \frac{1}{x} + 2x \ln x$
   (k) $(\sin x + \cos x)e^x \sec^2 x$ (l) $\frac{\sin x - x \ln x \cos x}{x \sin^2 x}$
   (m) $(8 - x^3)e^{-x}$ (n) $\frac{3x(x^2+2x+2)\ln x - x^3 - 3x^2 - 6x + 2}{x(\ln x)^2}$
2. (a) 1 (b) $\frac{1}{2}$
3. (a) $2 - e^x$ (b) $\frac{1}{x^2} - \frac{2}{x^3}$

Exercise 9.1

1. (a) $12x^{11}(2x+5)^5(3x+5)$ (b) $\frac{2}{\sqrt{9+4x}}$
   (c) $\frac{4x}{3\sqrt{4x^2+5}}$ (d) $-5\sin 5x$
   (e) $6\cos(6x-7)$ (f) $4\sin^3 x \cos x$
   (g) $-30\cos^4(6x-7)\sin(6x-7)$ (h) $12x\cos 3x + 4\sin 3x$
   (i) $24x^2 \sec^2(8x^3+1)$ (j) $\frac{\sec^2 x}{x+2} - \frac{\tan x}{(x+2)^2}$
   (k) $6e^{3x} + 4$ (l) $(1+2x^2)e^{x^2}$
   (m) $\frac{2x - x^2}{e^x}$ (n) $\frac{1}{x}$
   (o) $\frac{-2}{5-2x}$ (p) $\frac{-2x}{1-x^2}$
   (q) $\frac{1}{2x+11}$ (r) $3 + 3\ln x$
   (s) $\frac{1}{x\ln x}$ (t) $\frac{e^{x^2}}{x} + 2xe^{x^2}\ln x$
   (u) $e^{\tan x}\sec^2 x$ (v) $e^x \sec^2(e^x)$
   (w) $5e^{5x}\cos(e^{5x})$ (x) $5e^{\sin 5x}\cos 5x$
   (y) $\frac{-8x\sin[\ln(4x^2+9)]}{4x^2+9}$ (z) $-8x\tan(4x^2+9)$
2. (a) $2^{x^2+2}x\ln 2$
   (b) $x^x(1+\ln x)$
   (c) $(\sin x)^{\cos x - 1}\cos^2 x - (\sin x)^{\cos x + 1}\ln(\sin x)$
   (d) $\frac{-16x(2x+1)(3x+4)^5}{(x^2+7)^9} + \frac{15(2x+1)(3x+4)^4}{(x^2+7)^8} + \frac{2(3x+4)^5}{(x^2+7)^8}$
3. (a) $0.4 + 0.002x$ ppm/thousand
   (b) 25 thousand/year
   (c) 30.8 ppm/year
4. (a) 24

Exercise 9.2

1. (a) $\dfrac{2x - y^2}{2xy + 1}$ (b) $\dfrac{x^2 - 2y}{2x - y^2}$

   (c) $\dfrac{5x^4 + 4y^3}{3y^2(5y^2 - 4x)}$ (d) $\dfrac{-\sin y}{2y + x\cos y}$

   (e) $\dfrac{\cos(x + y) + y\sin x}{\cos x - \cos(x + y)}$ (f) $\dfrac{2x}{e^y - 2y}$

   (g) $\dfrac{y(3x^2 - \ln y)}{x + 2y^2}$ (h) $\dfrac{1 - (x + y)(e^y + y\cos x)}{(x + y)(xe^y + \sin x) - 1}$

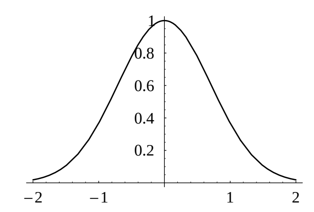
2. (a) $\dfrac{1}{8}$ (b) $\dfrac{1}{5}$ (c) -1

   (d) $\dfrac{1}{3}$ (e) -1 (f) 0

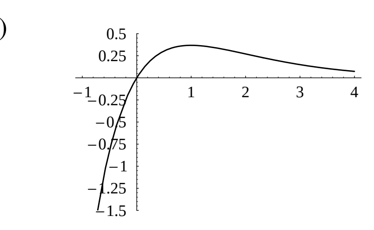
3. Decrease at the rate of 7.5 units per second.
4. Increase at the rate of 1280π cm^3 per minute.
5. $\dfrac{2}{\sqrt{5}}$ m/s

Exercise 9.3

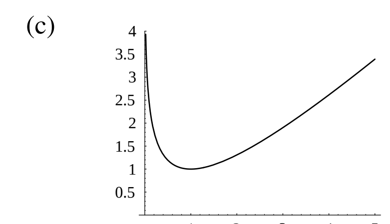
1. (a)



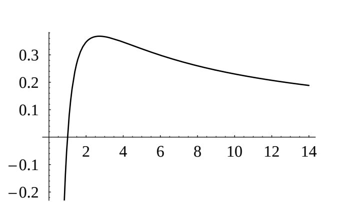
(b)



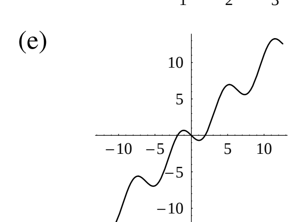
(c)



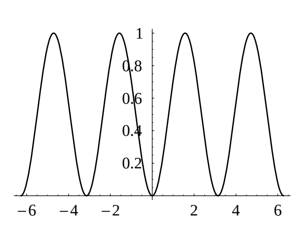
(d)



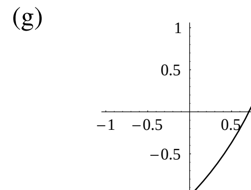
(e)



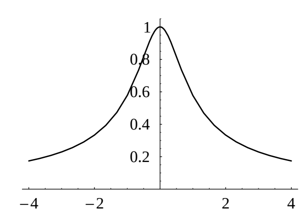
(f)



(g)



(h)



Exercise 9.4

1. $\sqrt{2e^{-1}}$
2. $q = 20$, price = 36.79, revenue = 735.76
3. (a) $t = 12$, $N = 50$

   (b) $t = 1$, $N \approx 81$

4. (a) 1 km from A
   (b) at A
5. $\frac{3}{2\sqrt{2}}$ m
6. $\frac{\pi}{3}$

Exercise 10.1

1. (a) $3\tan x + C$ (b) $2e^x + \sin x + C$
   (c) $2x + 3\ln|x| + C$ (d) $x - \frac{1}{x} + 2\ln|x| + C$
2. (a) 1 (b) $2e - \frac{2}{e}$
   (c) $\frac{1}{e} - \frac{1}{e^4} - \ln 4$ (d) $\ln 4 - 1$
3. (a) 2 (b) $1 + \ln 2$ (c) $2 - e^{-1} - e^{-2}$

Exercise 10.2

1. (a) $\frac{1}{10}(x^2+1)^{10} + C$ (b) $\frac{2}{15}(x^5+6)^{\frac{3}{2}} + C$
   (c) $-\frac{1}{2}\cos x^2 + C$ (d) $-\frac{1}{3}\cos^3 x + C$
   (e) $e^{x^2} + C$ (f) $\tan(e^x) + C$
   (g) $-\frac{1}{2}e^{-x^2+1} + C$ (h) $\frac{1}{3}e^{x^3-1} + C$
   (i) $\frac{1}{2}\ln(x^2+1) + C$ (j) $\cos\frac{1}{x} + C$
   (k) $\frac{1}{16}(x^2+2x+3)^8 + C$ (l) $\frac{-1}{40(x^4+2x^2+3)^{10}} + C$
   (m) $\frac{1}{5}(e^x - 3x)^5 + C$ (n) $2e^{\sqrt{x}} + C$
   (o) $\frac{1}{2}(\ln|x+1|)^2 + C$ (p) $\frac{1}{2}\ln|2x+7| + C$
   (q) $\frac{1}{16}(x+1)^{16} + C$ (r) $\frac{2}{3}(x-2)\sqrt{x+1} + C$
   (s) $\frac{1}{17}(x+1)^{17} - \frac{1}{16}(x+1)^{16} + C$ (t) $e^{x+\frac{1}{x}} + C$
2. (a) $\frac{21}{4}$ (b) $\frac{1}{2}(e^2 - e)$ (c) $\frac{1}{3}\sin 1$
   (d) $\frac{1}{2}(\ln 7 - \ln 3)$ (e) $e - 1$ (f) 2
   (g) $\frac{-122}{5}$ (h) $\ln 2$ (i) $\frac{1}{72}$
   (j) $\frac{40}{3}$
3. (a) 1 (b) $\frac{1}{2}(e - 2)$

Appendix B

Supplementary Notes

B.1 Mathematical Induction

Consider the following formula

$$1 + 2 + 3 + \cdots + n = \frac{n(n+1)}{2}. \tag{B.1.1}$$

Note that (B.1.1) involves n (positive integer). If we denote the equality by $P(n)$, then the statement "$P(n)$ *is true for all positive integers* n" means that "$P(1)$ *is true,* $P(2)$ *is true,* $P(3)$ *is true, and so on*". One way to proof this is to use mathematical induction.

Principle of Mathematical Induction Let $P(n)$ be a statement involving a positive integer variable n. Suppose that the following two conditions hold:

(I) $P(1)$ is true;

(II) $P(k+1)$ is true whenever $P(k)$ is true.

Then $P(n)$ is true for all positive integers n.

The above principle is easy to understand because (I) together with (II) implies that $P(2)$ is true which in turn together with (II) implies that $P(3)$ is true, and so on. To prove the principle rigorously, we have to use a property of natural numbers, namely, the *well ordering property*. This concept is discussed in more advanced books on sets.

Example Use mathematical induction to show that (B.1.1) is true for all positive integers n.

Proof Denote (B.1.1) by $P(n)$.

(I) When $n = 1$, we have

$$L.S. = 1 \qquad \text{and} \qquad R.S. = \frac{(1)(2)}{2} = 1.$$

Therefore $P(1)$ is true.

(II) Suppose that $P(k)$ is true, that is,

$$1 + 2 + 3 + \cdots + k = \frac{k(k+1)}{2}. \tag{B.1.2}$$

Then we have

$$\begin{aligned}
1+2+3+\cdots+k+(k+1) &= \frac{k(k+1)}{2}+(k+1) && \text{By (B.1.2)}\\
&= \frac{k(k+1)+2(k+1)}{2}\\
&= \frac{(k+1)(k+2)}{2}\\
&= \frac{(k+1)[(k+1)+1]}{2}
\end{aligned}$$

that is, $P(k+1)$ is true.

Thus by the Principle of Mathematical Induction, $P(n)$ is true for all positive integers n. □

Example Use mathematical induction to show that the following is true for all positive integers n:

$$1^2+2^2+3^2+\cdots+n^2=\frac{n(n+1)(2n+1)}{6}. \tag{B.1.3}$$

Proof Denote (B.1.3) by $P(n)$.

(I) When $n=1$, we have

$$L.S.=1 \quad \text{and} \quad R.S.=\frac{(1)(2)(3)}{6}=1.$$

Therefore $P(1)$ is true.

(II) Suppose that $P(k)$ is true, that is,

$$1^2+2^2+3^2+\cdots+k^2=\frac{k(k+1)(2k+1)}{6}. \tag{B.1.4}$$

Then we have

$$\begin{aligned}
1^2+2^2+3^2+\cdots+k^2+(k+1)^2 &= \frac{k(k+1)(2k+1)}{6}+(k+1)^2 && \text{By (B.1.4)}\\
&= \frac{k(k+1)(2k+1)+6(k+1)^2}{6}\\
&= \frac{(k+1)[k(2k+1)+6(k+1)]}{6}\\
&= \frac{(k+1)(2k^2+7k+6)}{6}\\
&= \frac{(k+1)(k+2)(2k+3)}{6}\\
&= \frac{(k+1)[(k+1)+1][(2(k+1)+1]}{6}
\end{aligned}$$

that is, $P(k+1)$ is true.

Thus by the Principle of Mathematical Induction, $P(n)$ is true for all positive integers n. □

Example Use mathematical induction (together with the product rule) to prove the power rule for positive integers n:

$$\frac{\mathrm{d}}{\mathrm{d}x}x^n=nx^{n-1} \tag{B.1.5}$$

Proof Denote (B.1.5) by $P(n)$.

(I) $P(1)$ is true since $\dfrac{\mathrm{d}}{\mathrm{d}x}x = 1 = x^0$ (by the convention for the function x^0).

(II) Suppose that $P(k)$ is true, that is,
$$\frac{\mathrm{d}}{\mathrm{d}x}x^k = kx^{k-1}. \tag{B.1.6}$$
Then we have
$$\begin{aligned}
\frac{\mathrm{d}}{\mathrm{d}x}x^{k+1} &= \frac{\mathrm{d}}{\mathrm{d}x}(x^k \cdot x) \\
&= x \cdot \frac{\mathrm{d}}{\mathrm{d}x}x^k + x^k \cdot \frac{\mathrm{d}}{\mathrm{d}x}x && \text{Product Rule} \\
&= x \cdot kx^{k-1} + x^k \cdot 1 && \text{By (B.1.6)} \\
&= kx^k + x^k \\
&= (k+1)x^{(k+1)-1}
\end{aligned}$$
that is, $P(k+1)$ is true.

Thus by Principle of Mathematical Induction, $P(n)$ is true for all positive integers n. □

B.2 Binomial Theorem

Before considering the Binomial Theorem, we introduce the notations $n!$ and $\binom{n}{k}$.

Consider a collection of three objects a, b and c. There are 6 *permutations* of the three objects:

$$\begin{array}{cc} abc & acb \\ bac & bca \\ cab & cba \end{array}$$

Instead of writing down all the permutations, we can find the number of permutations as follows.

- Note that there are 3 choices for the first object. Once the first one is fixed, there are two choices for the second object. Once the first and second objects are fixed, the third one is determined. Thus the number of permutations is $3 \times 2 \times 1$.

In general, given a collection of n objects, the number of permutations of the n objects is
$$n \cdot (n-1) \cdot (n-2) \cdots 3 \cdot 2 \cdot 1.$$

Notation Let n be a positive integer. We denote $n!$ (read "*n factorial*") to be the product of the first n positive integers, that is,
$$n! = n(n-1)(n-2) \cdots 3 \cdot 2 \cdot 1$$
By convention, $0!$ is defined to be 1.

Example $5! = 5 \times 4 \times 3 \times 2 = 120$.

Let A be the set having 5 elements a, b, c, d and e. The 2-element subsets of A are
$$\{a,b\} \quad \{a,c\} \quad \{a,d\} \quad \{a,e\} \quad \{b,c\} \quad \{b,d\} \quad \{b,e\} \quad \{c,d\} \quad \{c,e\} \quad \text{and} \quad \{d,e\}.$$

Instead of writing down all the 2-element subsets, we can find the number of such subsets as follows.

- First we consider ordered pairs of distinct elements of A:

$$(a,b),\ (a,c),\ \ldots\ (a,e), \qquad (b,a),\ \ldots\ (b,e), \qquad (e,a),\ \ldots\ (e,d).$$

  There are 5 choices for the first element and 4 choices for the second element.

- However for sets, $\{a,b\}$ and $\{b,a\}$ are equal. So the number of 2-element subsets of A is

$$5 \times 4 \div 2.$$

In general, given a set A with n elements, the number of k-element subsets of A can be found as follows:

- First we consider ordered k-tuples of distinct elements of A:

$$(x_1, \ldots, x_k) \qquad \text{where } x_i \in A \text{ and } x_i \neq x_j \text{ if } i \neq j.$$

  There are n choices for the first element, $(n-1)$ choices for the second element, $(n-2)$ choices for the third element and so on such that the number of choices for the k-th element is $(n-k+1)$. Hence the number of ordered k-tuples of distinct elements of A is

$$n(n-1)(n-2)\cdots(n-k+1).$$

- However for sets, $\{x_1, x_2, x_3, \ldots, x_k\}$ and $\{x_2, x_1, x_3, \ldots, x_k\}$ are equal. In fact, given an ordered k-tuple $\{x_1, x_2, \ldots, x_k\}$ of distinct elements of A, the sets formed by taking any permutation of the elements $x_1, \ldots, x_k$ are the same. So the number of k-element subsets of A is

$$n(n-1)(n-2)\cdots(n-k+1) \div k!.$$

Note that
$$\begin{aligned}\frac{n(n-1)(n-2)\cdots(n-k+1)}{k!} &= \frac{n(n-1)(n-2)\cdots(n-k+1)\times(n-k)(n-k-1)\cdots 2\cdot 1}{k!\times(n-k)(n-k-1)\cdots 2\cdot 1}\\ &= \frac{n!}{k!\,(n-k)!}\end{aligned}$$

Notation Let n be a natural number (positive integer or zero) and let k be a natural number not greater than n. We denote

$$\binom{n}{k} = \frac{n!}{k!\,(n-k)!}$$

Example $\binom{5}{3} = \frac{5!}{3!\cdot 2!} = \frac{5\cdot 4}{2} = 10$ and $\binom{5}{2} = \frac{5!}{2!\cdot 3!} = 10.$

Remark $\binom{n}{k}$ is the number of *combinations* that k objects can be chosen from a collection of n objects.

Example

$$\begin{aligned}\binom{n}{0} &= \frac{n!}{0!\,n!} = 1 & \binom{n}{n} &= \frac{n!}{n!\,(n-n)!} = 1\\ \binom{n}{1} &= \frac{n!}{1!\,(n-1)!} = n & \binom{n}{n-1} &= \frac{n!}{(n-1)!\,(n-(n-1))!} = n\\ \binom{n}{2} &= \frac{n!}{2!\,(n-2)!} = \frac{n\,(n-1)}{2} & \binom{n}{n-2} &= \frac{n!}{(n-2)!\,(n-(n-2))!} = \frac{n\,(n-1)}{2}\end{aligned}$$

Note $\binom{n}{k} = \binom{n}{n-k}$

The following result will be used in the proof of the Binomial Theorem.

Lemma B.2.1 *Let n be a positive integer and let k be a positive integer not greater than n. Then we have*

$$\binom{n}{k} + \binom{n}{k-1} = \binom{n+1}{k}$$

Proof

$$\begin{aligned}
\binom{n}{k} + \binom{n}{k-1} &= \frac{n!}{k!\,(n-k)!} + \frac{n!}{(k-1)!\,(n-(k-1))!} \\
&= \frac{n!}{k!\,(n-k)!} + \frac{n!}{(k-1)!\,(n-k+1)!} \\
&= \frac{n!\cdot(n-k+1) + n!\cdot k}{k!\,(n-k+1)!} \\
&= \frac{n!\cdot(n+1)}{k!\,(n-k+1)!} \\
&= \frac{(n+1)!}{k!\,(n+1-k)!} \\
&= \binom{n+1}{k}
\end{aligned}$$

□

Binomial Theorem *Let a and b be real numbers. Then for every positive integer n, we have*

$$(a+b)^n = \sum_{k=0}^{n} \binom{n}{k} a^{n-k} b^k \tag{B.2.1}$$

where by convention 0^0 means 1 if $a = 0$ or $b = 0$.

Note

$$\begin{aligned}
\sum_{k=0}^{n} \binom{n}{k} a^{n-k} b^k &= \binom{n}{0} a^n b^0 + \binom{n}{1} a^{n-1} b^1 + \binom{n}{2} a^{n-2} b^2 + \cdots + \binom{n}{n-2} a^2 b^{n-2} + \binom{n}{n-1} a^1 b^{n-1} + \binom{n}{n} a^0 b^n \\
&= a^n + na^{n-1}b + \frac{n(n-1)}{2} a^{n-2} b^2 + \cdots + \frac{n(n-1)}{2} a^2 b^{n-2} + nab^{n-1} + b^n
\end{aligned}$$

Proof Denote (B.2.1) by $P(n)$.

(I) When $n = 1$, we have

$$L.S. = (a+b)^1 = a + b \qquad \text{and} \qquad R.S. = \sum_{k=0}^{1} \binom{1}{k} a^{1-k} b^k = a^1 b^0 + a^0 b^1 = a + b.$$

Thus $L.S. = R.S.$ and so $P(1)$ is true.

(II) Suppose that $P(N)$ is true, that is

$$(a+b)^N = \sum_{k=0}^{N} \binom{N}{k} a^{N-k} b^k. \tag{B.2.2}$$

Then we have

$$\begin{aligned}
(a+b)^{N+1} &= (a+b)^N \cdot (a+b) \\
&= (a+b)\sum_{k=0}^{N}\binom{N}{k}a^{N-k}b^k && \text{By B.2.2} \\
&= \sum_{k=0}^{N}\binom{N}{k}a^{N-k+1}b^k + \sum_{k=0}^{N}\binom{N}{k}a^{N-k}b^{k+1} \\
&= \sum_{k=0}^{N}\binom{N}{k}a^{N-k+1}b^k + \sum_{k=1}^{N+1}\binom{N}{k-1}a^{N-(k-1)}b^{(k-1)+1} && \text{Replace } k \text{ by } k-1 \text{ and shift summation index} \\
&= a^{N+1} + \sum_{k=1}^{N}\left[\binom{N}{k}+\binom{N}{k-1}\right]a^{N-k+1}b^k + b^{N+1} \\
&= a^{N+1} + \sum_{k=1}^{N}\binom{N+1}{k}a^{N-k+1}b^k + b^{N+1} && \text{By Lemma B.2.1} \\
&= \sum_{k=0}^{N+1}\binom{N+1}{k}a^{N+1-k}b^k
\end{aligned}$$

That is, $P(N+1)$ is true.

Hence by the Principle of Mathematical Induction, $P(n)$ is true for all positive integers n. □

B.3 Mean Value Theorem

Mean Value Theorem *Let f be a function that is continuous on $[a, b]$ and differentiable on (a, b), where $a, b \in \mathbb{R}$ and $a < b$. Then there exists $\xi \in (a, b)$ such that*

$$f'(\xi) = \frac{f(b) - f(a)}{b - a}.$$

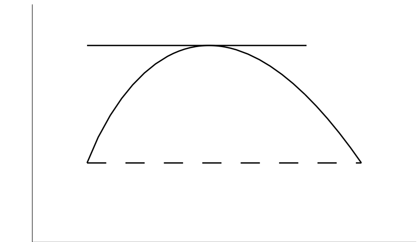
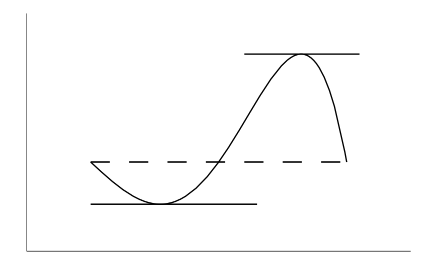
Explanation The conclusion means that there always exists (at least) a point $C = (\xi, f(\xi))$ on the graph of f, between $A = (a, f(a))$ and $B = (b, f(b))$, such that the slope at C equals to the slope of the line AB.



Before giving the proof for the Mean Value Theorem, we prove a special case of the result.

Rolle's Theorem *Let g be a function that is continuous on $[a, b]$ and differentiable on (a, b), where $a, b \in \mathbb{R}$ and $a < b$. Suppose that $g(a) = g(b)$. Then there exists $\xi \in (a, b)$ such that*

$$g'(\xi) = 0.$$

Proof We may assume that $g(a) = g(b) = 0$; otherwise, we can replace g by g_1 where $g_1(x) = g(x) - g(a)$.

(*Case 1*) g is identically zero on $[a, b]$

In this case, $g'(x) = 0$ for all $x \in (a, b)$. So any $\xi \in (a, b)$ satisfies the requirement.

(*Case 2*) g is not identically zero on $[a, b]$

By the Extreme Value Theorem, g attains its maximum and minimum in $[a, b]$, that is, there exists ξ_1 and ξ_2 in $[a, b]$ such that

$$g(\xi_1) \leq g(x) \leq g(\xi_2) \quad \text{for all } x \in [a, b].$$

Since g is not identically zero on $[a, b]$, it follows that at least one of ξ_1, ξ_2 is not an endpoint of $[a, b]$. Hence there exists $\xi \in (a, b)$ such that g has a local maximum or minimum at ξ. Therefore, by Theorem 5.1.3, we have $g'(\xi) = 0$.

□

Proof of the Mean Value Theorem Instead of working on f, we construct an auxiliary function g so that Rolle's Theorem can be applied to g and the conclusion for g is what we want for f.

Let $g : [a, b] \longrightarrow \mathbb{R}$ be the function defined by

$$g(x) = f(x) - f(a) - \frac{f(b) - f(a)}{b - a}(x - a) \qquad \text{for } x \in [a, b].$$

Note that g is continuous on $[a, b]$ and differentiable on (a, b) with

$$g'(x) = f'(x) - \frac{f(b) - f(a)}{b - a} \qquad \text{for } x \in (a, b).$$

Moreover, we have $g(b) = g(a) = 0$. Hence by Rolle's Theorem, there exists $\xi \in (a, b)$ such that $g'(\xi) = 0$, or equivalently that

$$f'(\xi) = \frac{f(b) - f(a)}{b - a}.$$

□

Remark If f is a function that is differentiable on an open interval I, then for every $x_1, x_2 \in I$ with $x_1 < x_2$, the function f is continuous on $[x_1, x_2]$ and differentiable on (x_1, x_2) and so we can apply the Mean Value Theorem to f with $a = x_1$ and $b = x_2$.

Below we apply the Mean Value to prove the following result which is Theorem 5.1.1.

Theorem B.3.1 *Let f be a function that is defined and differentiable on an open interval (a, b).*

(1) *If $f'(x) > 0$ for all $x \in (a, b)$, then f is increasing on (a, b).*

(2) *If $f'(x) < 0$ for all $x \in (a, b)$, then f is decreasing on (a, b).*

(3) *If $f'(x) = 0$ for all $x \in (a, b)$, then f is constant on (a, b), that is, $f(x_1) = f(x_2)$ for all $x_1, x_2 \in (a, b)$, or equivalently, there exists a real number c such that $f(x) = c$ for all $x \in (a, b)$.*

Proof We give the proof for (1) and (3). The proof for (2) is similar to that for (1). Alternatively, to prove (2), we may apply (1) to the function $-f$.

(1) Suppose that $f'(x) > 0$ for all $x \in (a, b)$. Let $x_1, x_2 \in (a, b)$ where $x_1 < x_2$. By the Mean Value Theorem, there exists $\xi \in (x_1, x_2) \subseteq (a, b)$ such that

$$f'(\xi) = \frac{f(x_2) - f(x_1)}{x_2 - x_1}.$$

Therefore, we have

$$f(x_2) - f(x_1) = (x_2 - x_1)f'(\xi) > 0,$$

which implies that $f(x_1) < f(x_2)$.

(3) Suppose that $f'(x) = 0$ for all $x \in (a, b)$. From the proof of (1), we see that for every pair of x_1, x_2 in (a, b), there exists $\xi \in (a, b)$ such that

$$f(x_2) - f(x_1) = (x_2 - x_1)f'(\xi) = 0.$$

which implies that $f(x_1) = f(x_2)$. Thus f is a constant function.

□

B.4 Fundamental Theorem of Calculus

Before giving the proof of the Fundamental Theorem of Calculus (Version 1), we need some preliminary results.

Lemma B.4.1 *Let f and g be functions that are continuous on a closed and bounded interval $[a, b]$. Suppose that $f(x) \leq g(x)$ for all $x \in [a, b]$. Then we have*

$$\int_a^b f(x)\,\mathrm{d}x \leq \int_a^b g(x)\,\mathrm{d}x.$$

Proof By definition, we have

$$\int_a^b (g - f)(x)\,\mathrm{d}x = \lim_{n\to\infty} \sum_{i=1}^{n} (g - f)(x_{i-1}) \cdot \frac{b - a}{n}, \quad \text{where } x_i = a + \frac{i}{n}(b - a) \text{ for } 0 \leq i \leq n.$$

The condition on f and g implies that each term in the above sum is non-negative. Hence we have

$$\int_a^b (g - f)(x)\,\mathrm{d}x \geq 0.$$

The required inequality then follows from Rules for Definite Integral (Int1) and (Int2). □

Corollary B.4.2 *Let f be a function that is continuous on a closed and bounded interval $[a, b]$. Suppose that m and M are real numbers such that $m \leq f(x) \leq M$ for all $x \in [a, b]$. Then we have*

$$m(b - a) \leq \int_a^b f(x)\,\mathrm{d}x \leq M(b - a).$$

Proof By Lemma B.4.1, we have

$$\int_a^b m\,\mathrm{d}x \leq \int_a^b f(x)\,\mathrm{d}x \leq \int_a^b M\,\mathrm{d}x.$$

The required inequalities then follow from Definite Integral for Constant. □

The next result is known as the *Mean Value Theorem for Definite Integral.*

Theorem B.4.3 *Let f be a function that is continuous on a closed and bounded interval $[a, b]$. Then there exists $\xi \in [a, b]$ such that*

$$\int_a^b f(x)\,\mathrm{d}x = f(\xi) \cdot (b - a).$$

Proof By the Extreme Value Theorem, there exist $x_1, x_2 \in [a, b]$ such that

$$f(x_1) \leq f(x) \leq f(x_2) \quad \text{for all } x \in [a, b].$$

By considering the constant functions $f(x_1)$ and $f(x_2)$ on the interval $[a, b]$ and applying Corollary B.4.2, we get

$$f(x_1) \cdot (b - a) \leq \int_a^b f(x)\,\mathrm{d}x \leq f(x_2) \cdot (b - a),$$

which yields

$$f(x_1) \leq \frac{\int_a^b f(x)\,\mathrm{d}x}{b - a} \leq f(x_2).$$

By the Intermediate Value Theorem, there exists ξ between x_1 and x_2 such that

$$f(\xi) = \frac{\int_a^b f(x)\,\mathrm{d}x}{b - a}.$$

Hence the required result follows. □

Fundamental Theorem of Calculus, Version 1 *Let f be a function that is continuous on a closed and bounded interval $[a, b]$. Let F be the function from $[a, b]$ into $\mathbb{R}$ defined by*

$$F(x) = \int_a^x f(t)\,\mathrm{d}t \qquad \text{for } a \leq t \leq b.$$

Then F is continuous on $[a, b]$ and differentiable on (a, b) with $F'(x) = f(x)$ for all $x \in (a, b)$.

Proof We divide the proof into two parts: continuity and differentiability.

(*Continuity*) By the Extreme Value Theorem, there exist real numbers m and M such that

$$m \leq f(x) \leq M \quad \text{for all } x \in [a, b]. \tag{B.4.1}$$

Let $x_1, x_2 \in [a, b]$ with $x_1 < x_2$. By the construction of F together with Rule for Definite Integral (Int3), we have

$$\begin{aligned} F(x_2) - F(x_1) &= \left(\int_a^{x_1} f(t)\,\mathrm{d}t + \int_{x_1}^{x_2} f(t)\,\mathrm{d}t\right) - \int_a^{x_1} f(t)\,\mathrm{d}t \\ &= \int_{x_1}^{x_2} f(t)\,\mathrm{d}t \end{aligned}$$

which, by (B.4.1) and Corollary B.4.2, yields

$$m(x_2 - x_1) \le F(x_2) - F(x_1) \le M(x_2 - x_1). \tag{B.4.2}$$

For every $\alpha \in (a, b)$, putting $x_1 = \alpha$ and $x_2 = x$, by (B.4.2) and the Sandwich Theorem, we see that

$$\lim_{x\to\alpha+} (F(x) - F(\alpha)) = 0,$$

that is, $\lim_{x\to\alpha+} F(x) = F(\alpha)$; similarly putting $x_2 = \alpha$ and $x_1 = x$, we see that

$$\lim_{x\to\alpha-} F(x) = F(\alpha);$$

hence we have $\lim_{x\to\alpha} F(x) = F(\alpha)$, that is, F is continuous at α. Similarly, the function F is left-continuous at a and right-continuous at b. Therefore F is continuous on $[a, b]$.

(*Differentiability*) Let $x \in (a, b)$. We want to show that

$$\lim_{h\to 0} \frac{F(x+h) - F(x)}{h} = f(x).$$

For this, we consider left-side and right-side limits.

For $h > 0$, by the construction of F together with Rule for Definite Integral (Int3), we have

$$F(x+h) - F(x) = \int_x^{x+h} f(t)\,\mathrm{d}t.$$

Hence by Theorem B.4.3, there exists $\xi_h \in [x, x+h]$ such that

$$F(x+h) - F(x) = f(\xi_h) \cdot \big((x+h) - x\big) = f(\xi_h) \cdot h.$$

Note that as h tends to 0 from the right, the number ξ_h tends to x (from the right). Therefore, we have

$$\lim_{h\to 0+} \frac{F(x+h) - F(x)}{h} = \lim_{h\to 0+} f(\xi_h) = f(x) \tag{B.4.3}$$

by the continuity of f.

For $h < 0$, by the construction of F together with Rule for Definite Integral (Int3), we have

$$F(x) - F(x+h) = \int_{x+h}^{x} f(t)\,\mathrm{d}t.$$

Hence by Theorem B.4.3, there exists $\xi_h \in [x+h, x]$ such that

$$F(x) - F(x+h) = f(\xi_h) \cdot \big(x - (x+h)\big) = f(\xi_h) \cdot (-h).$$

Note that as h tends to 0 from the left, the number ξ_h tends to x (from the left). Therefore, we have

$$\lim_{h\to 0-} \frac{F(x+h) - F(x)}{h} = \lim_{h\to 0-} f(\xi_h) = f(x) \tag{B.4.4}$$

by the continuity of f.

Combining (B.4.3) and (B.4.4), we get

$$\lim_{h\to 0} \frac{F(x+h) - F(x)}{h} = f(x)$$

as required.

□

Remark Rule for Definite Integral (Int3) can be proved using a property of continuous functions on closed and bounded interval, namely, *uniform continuity*. This concept is discussed in more advanced calculus or analysis.

Index